

WHO steps closer to its responsibilities

Developments in biomedical research give rise to ethical dilemmas and public controversy around the world. New guidelines reflect the WHO's need to strengthen its role in helping governments and others address the issues.

"Hurried and premature legislation in the rapidly evolving field of genetics can be counterproductive. Legislation and guidelines should be based on a full and sound scientific and ethical assessment of the techniques concerned." Ironically, this sensible recommendation comes from draft bioethics guidelines being prepared by the World Health Organization (WHO) — the very agency that, in response to the cloning of Dolly the lamb in 1997, issued hasty comment that human reproductive cloning was "ethically unacceptable and contrary to human integrity and morality".

Worryingly, in endorsing this knee-jerk response, the WHO failed to provide enlightened discussion of the many complex issues raised by cloning, and instead meekly bowed to public and political pressure. Indeed, in its haste to please, the WHO assembly ignored the warning of its own working group on cloning: that the massive political riposte to Dolly smacked of "moral panic" rather than considered deliberation of the issues involved.

The authors of the draft WHO guidelines (see page 179), in an implicit disavowal of the WHO's superficial treatment of the issue, politely suggest that "elaboration of the ethical, scientific, social and legal considerations that are the basis of this call for the prohibition of reproductive cloning should continue". That such thinly veiled criticism should survive the agency's potent internal censoring procedures gives grounds for optimism that the WHO is now ready to play a more considered role in the international bioethics debate.

That is eminently desirable. Although both the United Nations Educational, Scientific and Cultural Organization (Unesco) and the Council of Europe have produced international texts on bioethics, the former principally addresses human-rights issues raised by advances in genetics, whereas the latter is inevitably restricted to one continent. Gro Harlem Brundtland's agency has no option but to position itself at the sharp end of public-health matters — and hence many of the issues raised by biomedical research.

The main motivation for the WHO setting in motion plans to

adopt a role in bioethics was not so much a measured response to the real challenges as a political reaction to criticisms of the agency's unpreparedness to deal with issues such as cloning, and apparent envy over the lead on bioethics taken by that other UN agency, Unesco. The outcome, however, is remarkably comprehensive. Although the draft guidelines avoid ruling on areas such as the morality of human embryo research — where international consensus is, in any case, impossible — they do capture an accumulating consensus in the international community on certain fundamental principles that should govern biomedical research and its applications.

To those unaccustomed to the inevitable vagueness of international texts, it would be too easy to dismiss the guidelines as yet another list of good intentions. That would be a mistake; bioethics is ultimately an arena in which scientists, politicians and other members of society have to communicate and seek common ground.

Furthermore, while the guidelines' contents may seem obvious to many biomedical researchers, they are not so for many politicians and other lay persons. Many in developing countries, in particular, testify to the usefulness of such texts as a model to which their politicians and administrations can be referred in shaping legislation.

As an intergovernmental organization, with a hot line to health ministries throughout the world, the WHO has a duty to provide leadership in bioethics. It is also well placed to put in perspective concerns about perceived new threats to 'human dignity' with existing threats in the real world — such as the scandalously low funding of tropical-disease research.

But if the WHO is to play a serious role in bioethics, Brundtland will need to move further than guidelines. The agency will need the funds and the manpower to produce well-researched, forward-looking and credible input into the international debate on bioethics and public health. Meanwhile, the exercise of making the guidelines publicly available for wide comment (see <http://helix.nature.com/wcs>) is a valuable step towards greater openness. □

A commissioner for our time

Out of political mayhem, an opportunity for a visionary.

While it may be unseemly to dance on the political grave of the recently departed, it is permissible to welcome a vacancy for a replacement. As of Monday this week (15 March), Europe's research commissioner Edith Cresson was gone, together with her 19 fellow commissioners, following an independent inquiry. Although the launch of the fifth Framework programme of research has been achieved, major challenges remain.

Two stand out above all: the need to boost scientific education and skills across the continent — which inevitably necessitates increased support for basic research, however technology-orientated it is made to appear; and the need to enhance the technological flexibility of the continent as manifested in emerging companies, growing in the

wake not only of Europe's giant success stories (Nokia, Ericsson, Glaxo-Wellcome, Siemens ...), but also of inward investors and fundamental academic research.

Leadership comes not only from an understanding of the opportunities and priorities, but also from the imagination displayed in addressing them, and the enthusiasm with which this imagination can be conveyed. Many have good memories of the depth, vision and political reach of Etienne Davignon, who founded the precompetitive programmes such as ESPRIT in the early 1980s in a bid to reassert Europe's technological prowess in the face of revitalized US competition. Is it too much to hope that a similar individual can now be found? □



Human Genome Project aims to finish 'working draft' next year

[WASHINGTON] For the second time in six months, leaders of the Human Genome Project have brought forward significantly a key deadline for sequencing the three billion bases of the human genetic code.

The National Human Genome Research Institute (NHGRI) announced on Monday (15 March) that scientists in the international sequencing effort are now aiming to complete a 'working draft' by spring next year — 18 months earlier than predicted in a five-year strategic plan released last autumn (see *Nature* 395, 207; 1998). The working draft will cover at least 90 per cent of the genome, says the institute.

The NHGRI and its British collaborator, the Wellcome Trust, also announced the completion of the genome project's three-year 'pilot' phase and the launch of the full-scale effort to finish sequencing the genome. To this end, they awarded \$81.6 million to three academic sequencing centres in the United States, and the Wellcome Trust announced that it is boosting funding of the Sanger Centre in Cambridge, England, by \$77 million over the next 12 months (see box).

The three US sequencing centres funded by NHGRI are at the Whitehead/Massachusetts Institute of Technology Center for Genome Research in Cambridge, Massachusetts; the Washington University School of Medicine in St Louis, Missouri; and Baylor College of Medicine in Houston, Texas.

These three centres are expected to produce about 60 per cent of the working draft, with the Sanger Centre producing about 33 per cent. The Joint Genome Institute of the US Department of Energy at Walnut Creek, California, will be responsible for the rest.

The project's leaders said that the efficiency and accuracy of sequencing achieved in the last year of the pilot phase allowed them confidently to bring forward the deadline for completion of the working draft. Meeting the goal will require a 2.5-fold increase in the current rate of sequence production, to 2.5 gigabases per year.

"This is not a blue sky, 'gee we hope we can do it' kind of massive increase in [sequence production]," says Francis Collins, NHGRI director, adding that the new timetable is "eminently doable".

The emphasis on the working draft flows from a hunger among researchers for useful sequence as quickly as they can get it. "It has become clear that the best service to the scientific community is to deliver the draft sequence rapidly, and then to circle back and perform in the course of another year-and-a-half at most the finishing on that sequence,"



Wellcome boost: the UK Sanger Centre (above) will receive \$77m more for its sequencing work.

says Eric Lander, the director of the Whitehead/MIT Center.

"In that way the vast majority of the data will become available rapidly. And by the time that the scientific community needs the precise finished product, it will be ready."

The government-funded researchers deny that the launch of a private effort to sequence the genome by Celera Genomics of Rockville, Maryland, had any influence on their timetable (see *Nature* 393, 101; 1998). "I would not want people to see this move as reactive to what's going on in the private sector at all," says Collins. "It's part of the plan. It has always been."

Others are less sure. Huntington Willard, chairman of the Department of Genetics at Case Western Reserve University in Cleve-

land, Ohio, suggests that it might be "kinder to state that the Celera announcement of a year ago woke up NHGRI to what could be done if the resources were made available".

Cross-checking by the centres on each others' work has shown that sequencers are meeting an international accuracy standard of no more than one error in 10,000 bases, with the largest centres achieving rates of fewer than one error in 100,000 bases. Costs have also been cut from \$2 per base pair three years ago to an average of 20 to 30 cents per base.

The centres have also shown a better than expected ability to increase their sequencing capacity using economies of scale. The Whitehead Centre, for instance, has increased sequence production 12-fold in the past year without any extra money.

Surmounting one technical hurdle at the end of last year was particularly helpful in allowing the sequencers to advance their timetable. A bottleneck in providing mapped clones to sequencing machines was removed by the fingerprinting of entire libraries of bacterial artificial chromosomes at Washington University.

This will allow the assembly of the whole genome in 400–500 kilobase contiguous stretches, or 'contigs', by June this year, drastically reducing the labour involved in preparing mapped clones for the machines.

Several other sequencing centres involved in the pilot project but not funded in this week's awards may receive funding in another round of awards that NHGRI will announce in May.

Meredith Wadman

Britain sought to speed-up sequencing efforts

[LONDON] Britain's Wellcome Trust has for some time been urging US funding agencies to bring forward the genome sequencing target date, according to John Sulston, director of the Sanger Centre. The trust has just announced a major increase in funding for the centre's sequencing efforts (see above).

Sulston says that the trust's desire for speed is principally based on concern to meet the desires of researchers and the needs of medical research. He adds that the capability of completing a draft of the

genome before schedule has existed for some time, but that US agencies have been keen to proceed with caution.

He denies that the new deadline reflects a desire to beat the 2001 target set by a parallel private-sector initiative to sequence the genome, headed by Craig Venter of Celera Genomics. "The scientific community want this data immediately, and we aim to give it to them," he says.

Sulston acknowledges that the Celera deadline has contributed to "raising the temperature" in the race to

obtain a complete sequence.

But he says that, ultimately, medical research has most to gain from the US/UK initiative, which is committed to releasing sequencing data as fast as possible into the public domain, rather than from Celera, whose position on patenting of gene sequences remains unclear.

Michael Morgan, chief executive of the Wellcome Trust's genome campus, says that Celera can still join the US/UK effort. "It would be of huge benefit if they do, as we'd get the job done that much quicker." **Ehsan Masood**

French ministries seek way out of lab supplies crisis

[PARIS] The French science ministry is engaging in administrative juggling with the finance ministry in a bid to overcome a crisis in the supply of laboratory reagents to research agencies.

The problems stem from the agencies' difficulty in meeting new procurement rules for public markets. The law now requires all orders above FF300,000 (US\$50,000) to be put out to competitive tender.

This sum is quickly reached by the big national agencies, which buy many items such as animals, scientific apparatus, biological reagents and journals in bulk. Companies submitting the winning bid become the sole source of supply to that institution for the year-long contract.

The finance ministry last year ended an exemption from these rules for researchers (see *Nature* 396, 297; 1998). As a result, agencies must now buy most supplies in competitive tenders at the start of the year; a delay in doing so at the biomedical agency Inserm has led to its established orders drying up before the new ones have come into effect.

The problem is less acute at the basic research agency, the Centre National de la Recherche Scientifique, which, according to one official at the science ministry, anticipated the difficulties better. As an emergency measure, the ministry last week negotiated bridging funding for orders from Inserm.

But the ministry is also pressing the finance ministry to take into account the specific needs of research agencies. It argues that the law is badly suited to research, as it is difficult to predict so far in advance the reagents that are needed.

The finance ministry also ruled that orders should specify the 'product' and not the 'supplier' — although any scientist knows that supposedly identical reagents can behave differently. One official at the finance ministry is unsympathetic to this claim. "If the products are different, researchers can buy from different suppliers, but it is up to them to show a technical imperative," he argues. "If one product works better than another, then they must be different."

One administrative manoeuvre under consideration — and which the science ministry hopes to have accepted within six months — would allow agencies to pass 'virtual contracts' with several suppliers for a particular product, with orders for specific products only being made as needed.

"The agencies would act like a super-market," explains a science ministry official. "They would order all the varieties of pasta on the market — and then allow scientists to choose between the brands." **Declan Butler**



Murky waters: millions of cubic metres of acidic water were pumped from the area affected by the spill.

Doñana clean-up 'left half the soil still contaminated'

[BARCELONA] The Spanish government was overoptimistic when it initially evaluated the consequences of a toxic spill from the Swedish company Boliden's pyrite mines into the Guadimar River and a small part of the Doñana National Park last April, according to the latest analysis.

The Higher Council of Scientific Research (CSIC), which was asked, in collaboration with universities, to monitor the clean-up operation, produced a map last week showing the current level of contamination in the 4,000 hectares of agricultural land that was covered by toxic sludge.

CSIC researchers estimate that, even after an intensive clean-up, half the soil contains acidity or heavy metals — especially arsenic and zinc — at concentrations that César Nombela, head of the CSIC, says "should still be called contaminated". Average pollution levels have been found to exceed 50 to 100 parts per million (p.p.m.) of arsenic, and 15 per cent of the area may contain more than 275 p.p.m. of arsenic. International standards suggest a maximum of 50 p.p.m.

CSIC's report conflicts with an announcement in February by the Ministry of Agriculture that 84 per cent of the agricultural soil affected by the toxic spill is now suitable for cultivation. According to CSIC, government scientists appear to have based their study on only three elements — zinc, copper and lead — and to have ignored arsenic.

The accident led to the removal of 4 million tons of waste, in the form of sludge, and 4 million cubic metres of acidic water.

The clean-up work ended in January, having cost US\$100 million. During the five months of intensive work, however, it was not possible to prevent the drying up and chemical oxidation of the sludge covering the river bed and 4,000 hectares of agricultural soils.

The oxidation increased the acidity of the soil, which was also penetrated by pyrite-

containing metals. The challenge now is to find ways of cleaning up the soil. Some chemical procedures are already being applied to reduce the acidity and to immobilize metals in an attempt to prevent them from penetrating into the groundwater.

CSIC's researchers are also considering using the technique of phytobioremediation to remove arsenic from the soil. This involves the use of certain plants, including *Brassica* species indigenous to Spain, that can take up arsenic in large quantities through the roots.

Despite the lack of experimental data on the usefulness of such methods, Nombela says CSIC's scientists "have demonstrated the absorptive capacity of arsenic by indigenous plants" and argues that this will "soon lead to field tests to evaluate practical applications". Work on achieving this by developing genetically modified plants, containing a bacterial gene, is said to be in its preliminary stages.

An international meeting in January in Seville discussed remediation schemes. But, rather than agreeing on the use of phytobioremediation, it was concluded that, partly as a result of the high concentrations of arsenic, there is no short- or medium-term solution to the pollution in Doñana's wetlands. Participants agreed that the only feasible strategy was to immobilize the toxic metals as an interim measure.

CSIC's strategy is controversial. Luis E. Santamaria-Galdón, an ecologist at the Netherlands Institute of Ecology, says he cannot understand why CSIC insists on using transgenic plants and microorganisms: "The problem is too urgent to start developing previously unknown techniques."

But Nombela argues that "the molecular analysis of genes and biochemical mechanisms of absorption of pollutants offers an obvious path for scientific exploration that may also contribute to the development of new technologies." **Xavier Bosch**

WHO's bioethics code likely to stir debate

[PARIS] A set of bioethics principles drawn up by the World Health Organization (WHO) are likely to form the basis of the first comprehensive international guidelines on the potential risks of developments in biology and medicine.

The draft guidelines are due to be presented to the general assembly of WHO in May. They are likely to prove controversial. For example, they strongly support the rights of populations that are the subject of genetic research to have an "equitable share in the fruits of research, and a financial stake in any profits on resulting products".

They demand that genetic research on populations should respect "group consent, confidentiality, and the protection of the group's identity, culture, reputation and traditional beliefs". And they call for all research results from human genome projects to be made public and their fruits distributed equally.

Patents should not be granted on genes *per se*, but only on products developed from them, according to the guidelines, which describe the genome as the "common heritage of humankind". The guidelines also call for a ban on the use of genetic information to refuse employment or insurance.

WHO's decision to draw up the guidelines was prompted by the political controversy following the cloning of Dolly the lamb in 1997. This exposed the agency's relative inactivity on bioethical issues, and led to a call from member states and the general



assembly for the organization to take a more active role.

In May 1998, Hiroshi Nakajima, the then director-general of WHO, commissioned recommendations for a strategy on the issue from Abdullah Daar, a transplant researcher and prominent bioethicist from Sultan Qaboos University in Oman, and Jean-François Mattei, a physician and member of the French parliament (Département Libérale, Bouches-du-Rhône). The draft guidelines were approved by the WHO executive board in January (full details are published exclusively on <http://helix.nature.com/wcs>).

The WHO is likely eventually to promulgate the guidelines as an international declaration. Although not legally binding, it would represent a strong political message as to what the organization's member states

consider to be a consensus on fundamental bioethical principles.

The guidelines will be presented to the general assembly by the current director-general, Gro Harlem Brundtland. Observers say the move is intended to give greater political prominence to the social impact of issues raised by advances in genetics.

"The text is extremely important," says Axel Kahn, a prominent French geneticist and member of the national ethics committee. "It commits not just the countries that have already posed the problem, but all the countries in the world, and requires them to debate these issues."

The United Nations approved a Universal Declaration on the Human Genome and Human Rights last year (see *Nature* 396, 297; 1998). But this is mainly concerned with human rights issues raised by genome research.

The WHO guidelines are more wide-ranging, says Kahn, and cover "most of the sensitive issues". Kahn applauds in particular the guidelines' affirmation that "hurried and premature legislation in the rapidly-evolving field of genetics can be counterproductive".

Regular review of bioethics legislation to take into account scientific and social change should be the norm, says the text. France, for example, has opted to review its bioethics legislation every five years.

But Kahn is one of several scientists who criticize parts of the text. He points out that a clause stating that germline gene therapy could eventually be acceptable fails to mention that there are few potential therapeutic indications, and fails to outlaw the technique as a means of 'enhancing' the genome.

Another controversial statement is that experiments in developing countries should require the approval of ethical bodies in those countries. Some would prefer it to state bluntly that scientists should not be allowed to carry out experiments in such countries that would not be permitted under the ethical rules they must follow at home.

Daar accepts such criticisms. "These are exactly the sort of comments that we want to take into account before the text goes before the WHO general assembly," he says.

One significant absence from the draft guidelines is any reference to human embryo research. Daar explains that it is impossible to achieve an international consensus on this issue, adding that "if we get a lot of comments saying we must address this issue, we will".

A report accompanying the guidelines that will be submitted by Daar and Mattei to the general assembly calls for WHO to set up a comprehensive unit to "explore the potential impact of new genetic knowledge and technology on health policies". **Declan Butler**

India falls into line over patents legislation

[NEW DELHI] After a delay of more than three years, the Indian parliament finally agreed last week to amend the country's 29-year-old patent law, bringing it in line with other members of the World Trade Organization.

The bill was passed, amid noisy protests and a walk-out by left-wing parties in the Lok Sabha (the lower house), five weeks before the 29 April deadline that had been set by the World Trade Organization's dispute settlement body. The Rajya Sabha (the upper house of parliament) had already approved the bill.

The passage of the Patents (Amendment) Bill 1999 in both houses of parliament means that companies will soon be able to obtain exclusive marketing rights (EMRs) for drugs and agrochemicals and to apply for product patents.

But included in the bill, at the insistence of the opposition Congress party – which voted with the ruling Bharatiya Janata Party in favour of the bill – is a ruling that no EMRs should be given to drugs based on indigenous medicines.

EMRs will be valid for five years, or until the patent application is rejected or granted, whichever comes first. The existing law in India forbids EMRs and provides for patents only on processes, not products.

Under the new law, product patents will be awarded from 1 January 2005 when a 'transition phase' comes to an end and the product patent regime becomes mandatory. "More than 3,000 applications are currently in the 'mailbox' for product patents," says industry minister Sikandar Bakht.

Bakht told parliament that, although companies will benefit from EMRs, the bill contained enough safeguards on pricing, compulsory licensing and powers to revoke EMR licences in the national interest.

Protests against the bill came mainly from communist members of the lower house, who argued that the amendments were "anti-national" and a "sell-out" to multinational companies that want to destroy India's indigenous drugs industry. But, despite the protests, the bill was passed by 231 votes to 55.

K. S. Jayaraman

Call for a lighter regulatory burden on NIH researchers

[WASHINGTON] The US National Institutes of Health (NIH) should reduce the burden of federal regulations facing the university-based and extramural scientists it funds, according to a report released last week.

The report, commissioned by the NIH at the request of Congress, calls for "concerted action and sustained attention," arguing that the rules are "in need of change, and in some cases, dramatic change".

According to the report, the burden of regulation should be reduced in the five areas it studied: conflict of interest, animal care and use, protection of human subjects, research integrity, and disposal of hazardous waste. It says that "all members of the community with whom these issues were discussed feel strongly, even passionately," about the need to address them, and quickly.

The report, *NIH Initiative to Reduce Regulatory Burden*, defines 'burden' as any aspect of a regulation that could be made more efficient without defeating its purpose. It was drawn up by John Mahoney, a management consultant and former NIH chief financial officer.

The report calls on the NIH to appoint a

single individual or office (separate from the offices that now enforce the regulations) to coordinate a response, guided by advisers from the research community.

Anthony Demsey, a senior policy adviser in NIH's Office of Extramural Research, calls the report "well considered" and says the NIH is "very supportive" of many of the recommendations, and wants to "address right away some of the very clear issues".

But "there are others that may be controversial," he adds. "Obviously we reserve the right to consider not just the report but the comments we get as to how we want to go about the implementation." Demsey concedes that some interested groups, such as the Humane Society, did not have active input into the report, but that the NIH will "welcome" their response.

Scientists are beset by problems in all five areas, the report notes: rigid regulations that do not allow flexible responses; rules imposed by multiple agencies with inconsistent requirements; oversight by non-scientists; and poor communications with government agencies.

But it warns against over-hasty attempts

to undo regulations, which often result from careful political compromises. It cautions that, without careful planning, amendments could leave researchers worse off.

Instead, it says, NIH should seek help from allies in Congress and the administration, such as the White House's Office of Science and Technology Policy, to gauge the climate for change. But it adds that now is as good a time as any to try for change.

Congress asked NIH to streamline regulations affecting its extramural scientists when it approved its funding in 1998. Mahoney, charged with consulting researchers, interviewed officials at seven federal agencies and departments, eight universities and 11 research organizations, and enlisted 50 extramural scientists and officials in working groups that studied the five areas.

Congressman Dan Miller (Republican, Florida), who inserted the provision in the spending law, says he is "pleased" that NIH has documented the "excessive waste" caused by regulations.

The report is an encyclopedia of the rules' apparent redundancy, rigidity and inappropriateness. For example, scientists must answer to three different sets of animal-care inspectors, with site visits at least three times a year, and consequent paperwork.

It also says that allowing an animal to be used in surgery only once may be increasing the number used in research. And it points out that the cage-size requirements of the US Department of Agriculture force scientists to keep social animals, particularly primates, in solitary conditions — a policy the National Research Council has called "indefensible".

The report also cites an "exceedingly and needlessly legalistic and complex" process by which institutions assure the NIH that they are complying with protections for human subjects.

In addition, the report calls for the Office of Research Integrity in the Department of Health and Human Services to be restricted to an educational function. It argues that the office's authority to reinvestigate scientific misconduct cases after the institution involved has already done so is duplicative and leads to "second-guessing" of the original findings, compromising their authority.

Regarding conflict of interest, the report argues that Public Health Service regulations trigger "an excessive number" of disclosures because of an unrealistically low equity threshold of \$10,000, or 5 per cent ownership of entities that could be affected by a scientist's research.

On hazardous waste disposal, the report says universities are burdened by regulations intended to apply to industrial waste, not the small volumes produced by laboratories.

The report can be found at <http://www.nih.gov/grants/policy/regulatoryburden/>. NIH is accepting comments for 30 days at regburd@od.nih.gov. **Meredith Wadman**

US quest for AIDS vaccine appoints a leader

[WASHINGTON] After an 18-month search, the National Institutes of Health (NIH) announced last week that Gary Nabel, a professor of internal medicine and biological chemistry at the University of Michigan at Ann Arbor, is to become director of its new Vaccine Research Center.

In a statement, President Bill Clinton said that Nabel will be an "incredible" asset in the search for an AIDS vaccine — the centre's top priority. When he announced the creation of the centre in May 1997, Clinton called for the development of an AIDS vaccine within a decade (see *Nature* **387**, 323; 1997).

The time it took to appoint a director has been criticized by activists such as Gregg Gonslaves, policy director at the Treatment Action Group, which lobbies for AIDS research. Also criticized is that Nabel, who developed the first gene



Nabel: wants a more scientific approach to vaccine research.

therapy for HIV infection and has developed DNA-based vaccines for melanoma and Ebola virus, is not an experienced vaccinologist or immunologist.

But NIH director Harold Varmus described Nabel as "a superb scientist" who is "remarkably well prepared" for the job. And David Baltimore, president of the California Institute of Technology and head of NIH's Aids Vaccine Research Committee — and under whom Nabel has worked — calls the appointment "great".

Nabel told *Nature* that he plans to develop a broad scientific understanding of immune responses to foreign proteins, allowing predictions of a candidate vaccine's effectiveness, as opposed to randomly picking candidates and hoping for the best. He also plans to stress the development of candidate vaccines to fill what is currently a nearly empty pipeline for phase-one clinical trials, and to work closely with industry.

His scientific priorities include assessing whether a successful vaccine can be developed using cytolytic T-cell immunity, and developing approaches that enhance the production of neutralizing antibodies.

Nabel says he believes that his strengths do lie in vaccinology and immunology. "What I have not done is work in industry on vaccine development." **M. W.**

Britain's 'high-tech' budget praised...

[LONDON] The British government's latest measures to boost knowledge-driven economic growth, announced in last week's budget, have been applauded by scientists, universities and investors (see *Nature* 398, 98; 1999).

Each of the measures announced last Tuesday addresses a particular stage of the commercialization of research (see table).

Most were inspired by a report published last year by a committee of scientists, entrepreneurs and investors, set up by the government and chaired by Sir Peter Williams, chairman of Oxford Instruments, to identify ways of improving the financing of high-technology companies.

The budget measures will be complemented in the coming weeks by a report from the Department of Trade and Industry outlining its plans for implementing its recent white paper on competitiveness (see *Nature* 396, 714; 1998).

The government last week also announced the winners of the first round of University Challenge, a competition for a £45 million (US\$73 million) seedcorn fund to help universities commercialize research. Britain's finance minister, Gordon Brown, increased

Budget highlights

* The starting rate of corporation tax will be halved next year to 10 per cent for companies whose annual profits do not exceed £10,000 (US\$16,260).

* A tax credit allowing small- and medium-sized companies not making profits to reduce the cost of research and development. Companies in profit will receive a 12.5 per cent R&D tax credit.

* A new £20 million venture capital fund in partnership with the private sector will provide finance for young, high-technology firms.

* A tax-free employee share-ownership scheme in which capital gains will also be tax free if kept in the scheme for three years. This is aimed at attracting top managers to small companies unable to pay large cash salaries.

* Corporation tax relief from next year for large companies that invest in small, risky ventures.

* £100 million for the Joint Infrastructure Fund. This was launched by the government and the Wellcome Trust last July as a £600 million fund for science laboratories, buildings and equipment.

* An additional £10 million for a second round of the University Challenge Fund, a seedcorn fund to commercialize university research.

the fund by £5 million so that promising initiatives that did not make the final shortlist would not have to find alternative sources.

Most of the winners were joint bids by groups of large research universities. But Lord Sainsbury, the science minister, said that the government is to repeat the competition with an additional £15 million available. He says he wants every university to be involved in commercializing its research.

Alan Wilson, vice-chancellor of the University of Leeds, welcomes the "incentives for technology transfer in the budget" but is

concerned that core university funding has only just kept pace with inflation.

Leeds is among a number of leading UK research universities to have enthusiastically embraced the government's technology-transfer initiatives. Together with the universities of Sheffield and York, Leeds has emerged as joint-highest recipient of University Challenge funding with £4.5 million.

Sir David Cooksey, a member of the Williams committee and chairman of the British Venture Capital Association, welcomed the measures in the budget. But he regrets the government's decision to omit a key recommendation from the Williams report: exemption from both capital-gains tax and corporation tax for large companies that invest in high-technology companies — known as corporate venturing.

The report said that such incentives could unlock vast amounts of capital for high-technology industry. Large investors such as insurance companies currently consider the high-technology sector too risky to invest in.

The government has, instead, decided to restrict tax incentives for corporate venturing to relief from corporation tax but not from capital-gains tax, although this has not been ruled out for the future.

Craig Pickering, another member of the Williams committee and a former senior civil servant at the Treasury, says that, in his experience, ideas take time to make the transition from a proposal to a change in the law. One of the reasons for this is that the government's taxation agency, the Inland Revenue, takes time to work out in detail the implications of any new proposals.

One Treasury official says that progress on further tax relief for corporate venturing is unlikely until the government is able to quantify the cost to the public purse of a more comprehensive tax rebate.

He says it is difficult to forecast the capital gains that a large company could make from high-technology investment. It is safer, he adds, for the government to set aside a fixed sum towards a venture-capital fund.

Ehsan Masood

... in the push for a knowledge-driven economy

[LONDON] The British government believes it is near the end of its checklist of measures to turn the country into a knowledge-driven economy. The central idea behind the initiatives is that richer countries with a good science base should seek to profit from knowledge, as their high labour costs make them uncompetitive in low-skilled areas of work.

This idea is backed up by a belief that knowledge-driven activities can make a substantial contribution to economic growth. Much of the government's vision is inspired by the success of the US information technology and biotechnology industries, which ministers and university vice-chancellors have been studying closely.

Science minister Lord Sainsbury, for example, is considering setting up regional clusters of technology businesses

similar to those in the United States. And the Committee of Vice-Chancellors and Principals organized a study visit to the United States, which resulted in a conference last month at which the former head of the technology transfer office at the Massachusetts Institute of Technology was flown over to advise UK academics.

But concerns are being expressed privately by companies involved in helping to get research to the marketplace. Many feel that the government is overemphasizing the role of university 'spin-out' companies; that its initiatives are too 'technology push' rather than 'demand pull'; and that it needs to understand better the differences between the experiences of Britain and the United States.

For example, much US high-technology industry is driven by scientists who return to universities after a

period in industry, not necessarily by scientists who set up spin-out companies, points out the head of a technology transfer office at a university in the north of England.

Second, the bulk of US high-technology investment comes from wealthy individuals, not large corporations.

And third, a shortage of effective managers is more of an obstacle to technology transfer in Britain than a shortage of funds.

This is the view of Ian Harvey, chairman of BTG plc, which specializes in the commercialization of publicly funded research. Harvey says that funds can always be found for a good concept with an effective management team.

"There is a danger that start-ups are becoming flavour of the month," he says. "But if you can't find people to manage them, you could be worse off." **E.M.**

Students win pay rise in Michigan strike

[WASHINGTON] The University of Michigan, where graduate students have been on strike for the past two weeks in protest over payments for teaching undergraduate courses, has agreed to increase such rates by 10.5 per cent over three years.

But still undecided is the issue of compensation for foreign teaching assistants and the English language courses they are required to take. The strike organizers say the dispute is part of an escalating movement among graduate students in universities across the United States.

The strike is the third major revolt by graduate students in recent years. Three years ago, students at Yale University went on strike over the treatment of low-paid campus employees. And last year, student teaching assistants took strike action at all eight campuses of the University of California over the right to form a union.

Most of the Michigan strikers are members of the Graduate Employee Organization, a union affiliated to the Federation of American Teachers. The union is in the process of organizing graduate students at 25 other US academic institutions.

The strike involves about two thirds of the 2,100 graduate student instructors on the Michigan campus. About half are science and engineering graduate students.

According to the university, they work about 20 hours a week, and earn an average of \$16.34 an hour, making Michigan one of the top US universities in terms of pay and conditions for the instructors. But the students say their monthly income of about \$1,133 is not enough to make ends meet.

Stephen Arellano, a union spokesman, says pressures on graduate students are intense. "The university likes to refer to us as professionals, but they don't treat us as such," he says. "We do a good job teaching and we should get compensated." The students were demanding a pay increase to \$1,400 a month.

Some see the events at Michigan as reflecting increasing tension in US research universities. "Graduate students see themselves today as part of a labour group," says one academic lobbyist in Washington. "They don't make enough money, they don't like their conditions, their fellowships [are increasingly] being taken away, they're in debt up to the hilt. This is going to grow."

The issue draws attention to the role of the Association of American Universities (AAU), whose members include most of the major research universities. Although a study last year confirmed the AAU's position that graduate students should primarily be considered as students, it said little about the financial and psychological strains on them.

AAU president Nils Hasselmo says he sees no need for the association to intervene in such conflicts, as universities should deal with militant graduate student unions case by case. But he says the AAU remains interested in the conditions of graduate education.

But others criticize the AAU for not taking a more active role in defending graduate students. One source affiliated to the National Academy of Sciences says: "These students don't want to be United Auto Workers; they want to be students and academics. But they don't want to be walked over, either."

Some predict that postdoctoral students, who are also under economic pressure, may take similar action. And there are also rumblings from adjunct professors, who make even less money than graduate assistants and work longer hours.

Wil Lepkowski

Japanese emissions plan under fire for relying on nuclear power

[TOKYO] Japan's latest plan to tackle global warming has been strongly criticized by environmentalists for its strong dependence on the use of nuclear energy to reduce the nation's emissions of greenhouse gases.

The basic plan on global warming, released last week by the Central Environmental Council of the Environment Agency, details measures for achieving the 6 per cent reduction in greenhouse-gas emissions from 1990 levels by 2008–12 that was agreed by Japan at the Kyoto climate conference in 1997.

The plan, now under consideration in parliament, calls for the promotion of nuclear energy to help reduce the emission of such gases. It proposes that 20 nuclear reactors should be built by 2010 to increase the electricity generated by nuclear power by 50 per cent over its 1997 level.

The plan also includes the development of alternative energy resources such as solar energy. But its main emphasis is on nuclear energy, although it says this should depend on there being adequate safety measures and proper management of radioactive waste.

The plan is based partly on a bill drawn up last year by the prefecture, which requires both central and prefectural governments to make plans to reduce greenhouse-gas emissions. But it derives mainly from the government's general policy outlines on global warming, also released last year,

which indicated that Japan's use of nuclear energy would need to be substantially increased to achieve the planned reduction (see *Nature* 393, 199 & 394, 3; 1998).

The first draft of the environment agency's plan focused on alternative energy resources and measures to reduce carbon dioxide emissions in industry, but did not mention nuclear energy. But this was changed after resistance from industrial lobbyists and the Ministry of International Trade and Industry (MITI).

Environmental groups have criticized



Tarnished: accidents at the Mihama power station and elsewhere have dented nuclear's image.

the plan as unrealistic, given the anti-nuclear mood in Japan following a series of accidents and cover-ups at nuclear facilities.

The electricity industry itself does not seem to want more nuclear reactors. Most electricity companies have frozen plans to build new reactors after strong opposition from local residents. Many are planning to increase the operation time of existing reactors and say that regular safety checks and repairs could double the lifespan of such reactors, normally 30 years.

Although central government cannot build nuclear reactors without the approval of prefectural governments, MITI says it will look again at the problem of radioactive disposal and renew its efforts to increase public support for more nuclear facilities.

But according to Kiko Forum, a federation of more than 100 environmental groups, Japan cannot meet its target for reducing emissions under the current plan without building 20 new reactors.

Mie Asaoka, director-general of Kiko Forum, says: "The problem lies in the fact that [the government] has been over-dependent on nuclear energy, and has failed to revise its long-term energy policy."

Asaoka adds that Japan's gross domestic product is declining, with little prospect for immediate recovery. "Perhaps we will not need all the energy the government claims we will," she says.

Asako Saegusa

European archive for mouse mutants is set to open at last

Alison Abbott

With the rapidly burgeoning numbers of mouse mutants being produced, Europe's need for a repository and distribution centre has become urgent. The announcement that its mouse archive is to open is thus welcome.

[ROME] The European Mouse Mutant Archive (EMMA), which has been under construction for the past three years, will open on 1 April, according to Peter Rigby, the co-chairman of its scientific policy committee.

EMMA, which has been built at Monterotondo near Rome with support from the European Commission, has been ready to accept mouse strains for several months. But its opening has been delayed because the committee, which oversees its activities, has not established all its principles of operation. There is also no full-time EMMA director.

The committee has now agreed to open the facility while detailed policy — for example on the criteria for selecting mouse strains for archiving — is being established.

At the same time, Europe's mouse-genetics community, which has complained about a lack of clarity in the committee's activities, is asking for wider representation of its members on the committee.

The committee appears to have been prompted into its announcement by last week's opening of the Adriano Buzzati Traverso international science campus, which EMMA shares with the Italian CNR Institute for Cell Biology, the European Molecular Biology Laboratory's (EMBL) Mouse Genetics Programme, and an outpost of the Trieste-based International Centre for Genetic Engineering and Biotechnology. EMMA's main site at Monterotondo will have 'nodes' in France, Britain, Portugal and Sweden.

Mouse-mutant archives, which serve as a library for mouse-mutant strains created or

discovered by researchers, are becoming particularly important now that the international mouse genome programme is under way. Mouse mutants will help scientists understand the functions of genes being identified.

Over the past year, EMMA's small staff has demonstrated the facility's ability to freeze and reconstitute viable mouse embryos and transport them across Europe. But a site visit last November by the European Commission resulted in a report expressing concern that EMMA's scientific policy committee appeared to meet rarely, and had been slow to develop policies for pricing, accepting and distributing strains, and for publishing information on its archives.

Committee members are reluctant to comment on such criticisms. But many scientists suggest that inadequate and uncertain funding has meant EMMA has been unable to attract a full-time on-site director who could have applied pressure for more frequent committee meetings.

Reflecting such concerns, a task force intended to track EMMA's progress was set up last October by the International Mammalian Genome Society (IMGS), after those attending its annual meeting wanted to know how EMMA would serve the community.

But Rudi Balling, director of mammalian genetics at the GSF Institute for Environment and Health, the Munich-based national research centre that organized the meeting, says "there was a feeling that the community must make sure the repository is a success, by giving it its full support".

The demand for a mouse repository and distribution centre, already high, is expected to grow rapidly in the next five years or so. "The entire mouse genome is likely to be roughly mapped by 2003," says Peter Gruss, director of molecular and cell biology at the Max Planck Institute for Biophysical Chemistry in Göttingen and the EMMA committee's other co-chairman. "So theoretically a mutant for every important gene would need to be archived in order to allow scientists to research their functions."

The research community needs EMMA to begin operating as soon as possible, says Philip Avner, a senior scientist at the Pasteur Institute in Paris and IMGS president. "As soon as it opens, EMMA will be flooded with applications from people wanting to park their transgenics."

"Laboratories which have been working for many years with mouse mutants want to be able to distribute their strains efficiently to other interested research labs," says Klaus Rajewsky, coordinator of the EMBL programme at Monterotondo.

His own laboratory at the University of Cologne receives several requests a week for strains of his gene-targeted mutant mice, he says. "But we cannot fulfil the requests simply because of the time it takes — the best part of a day for each pair of mice."

Pressure will only increase as large-scale chemical mouse-mutagenesis screens start to produce large numbers of important strains. These screens identify interesting characteristics created when mice are administered a mutagen.

The conclusions of the IMGS task force echo the commission's concerns at the delays in policy development. "The EMMA committee must also be as responsive to the community as possible," says Avner. "So it must be expanded to include large-scale mutagenesis scientists and classical geneticists."

Rigby, who is the director of the Institute for Cancer Research in London, says the committee will meet again as soon as possible. Top of its agenda will be discussions of acceptance criteria. But these will not be too rigid, he says, as EMMA must be able to respond to scientific needs and developments: "We will have to learn by doing." The committee will also discuss its own expansion, as Rigby agrees that it must encompass the new user communities.

Finding funds to pay for a director is a clear priority, says Rigby. EMMA's two current EC grants expire this year. Once the next round of funding is secure, he says, the committee can look at how to get more 'long-term' financing for the facility.

The commission itself is reluctant to create and support permanent institutions. But, says Balling, "there is complete consensus in the mouse genetics community that EMMA should somehow be institutionalized; it is ridiculous to create such a fundamental facility on short-term money." □



EMMA: faces concerns that its opening may have been delayed by uncertainty over funding prospects.

Marijuana should be tested in clinical trials, says US report

[WASHINGTON] A report from the US Institute of Medicine has found that marijuana is potentially effective in treating pain, nausea, AIDS-induced anorexia and other symptoms. It says the drug should be tested rigorously in clinical trials.

But its potential value as medicine "is undermined by the fact that patients must inhale harmful smoke," says Stanley Watson, co-principal investigator of the study, and co-director and research scientist at the Mental Health Research Institute, University of Michigan at Ann Arbor.

The study urged scientists to develop less harmful delivery systems, and concluded that the use of marijuana should be confined to patients with terminal illness and those with debilitating disease who have failed to respond to other medicines.

Voyage of collaboration for Middle East geologists

[MUNICH] Geologists and environmental scientists from Germany, Jordan, Egypt and Saudi Arabia put to sea last week for a joint geoscientific research expedition in the Red

Sea on board the German research vessel *Meteor*. The scientists will take sediment cores from the sea floor to reconstruct the climate history of the past 10,000 years.

The cruise is funded by the German research council, *Deutsche Forschungsgemeinschaft*, which approved a research cooperation agreement between Germany and the Middle East in 1996 (see *Nature* 379, 577; 1996). Earlier this year, the vessel hosted marine biologists from Israel and Palestine for oceanographic research in the Gulf of Aqaba and the northern Red Sea. The research council hopes that promoting collaboration between scientists will support the Middle East peace process.

Geneticists 'must work with social sciences'

[TOKYO] The bioscience panel of Japan's Science Council, an advisory body to the education minister, released an interim report on human genome research last week calling for a new approach that integrates genetics with disciplines such as information science and social sciences. The report also calls for the selection of a specific organism other than humans for carrying out intensive genome analysis.

Citing Japan's slow progress in the Human Genome Project, the panel says that Japanese

researchers will have to analyse 100 megabases of the human genome, 50 to 100 megabases of microbial genomes and 100 megabases of plant and animal genomes each year in order to catch up with the speed of researchers in the United States and Europe.

Nuclear physicist wins prize for religion

[BOSTON] Ian Barbour, a nuclear physicist and theologian, has won the 1999 Templeton Prize for Progress in Religion for his efforts to bring science and religion "together to the benefit of both disciplines". The award was announced last week at the United Nations in New York, and Barbour will receive the US\$1.24 million prize in London on 11 May.

Barbour earned a PhD in physics from the University of Chicago in 1949, working under the direction of Enrico Fermi. For more than three decades, Barbour has explored the interface between science and religion, addressing the ethical and moral issues raised by technological advances. Barbour has pledged to donate \$1 million of his prize money to the Center for Theology and the Natural Sciences in Berkeley, California. The award scheme was launched in 1972 by global investor Sir John Templeton.

Royal Society to review data in row over GMOs

[LONDON] Britain's Royal Society is to set up an expert group to assess the research at the centre of last month's media panic over genetically modified foods. Sir Aaron Klug, the society's president, said the group will review claims from Arpad Pusztai that rats fed with potatoes modified with a lectin gene suffered ill health. Pusztai was formerly a senior researcher at the Rowett Research Institute in Aberdeen.

"Premature, partial or selective release, or misinterpretation of unsubstantiated research only serves to mislead the general public in a complex area," said Klug. "Rational debate based on rigorously reviewed data is essential." The review is expected to be completed by late May.

The society's move comes amid reports that the government and industry in Britain are considering a three-year voluntary moratorium on the commercial planting of genetically modified crops.

Anniversary nostalgia hits Russian academy

[MOSCOW] The 275th jubilee of the Russian Academy of Sciences, to be commemorated in June, should be used to create a "united

scientific space all over the Commonwealth of Independent States," said Russian prime minister Yevgeny Primakov last week. He was addressing a meeting of the state committee organizing the celebration.

Primakov's proposal, supported by academy president Yuri Osipov, reflects a general mood of nostalgia at the higher echelons of power in Russia. In the former Soviet Union, the central Soviet Academy of Sciences was responsible for all the regional republican academies, and this model — relaxed after the fall of communism — appears attractive to the present science leadership in Moscow.

Scientists 'should take oath of ethics'

[LONDON] Scientists should consider adopting a pledge similar to the Hippocratic oath, committing themselves to high ethical standards, rigorous quality control of their findings, open access to their knowledge, and public accountability.

These are some of the conclusions of a draft declaration of science policymakers from Arab countries who met in Beirut last week to finalize the regional preparations for the World Conference on Science, which is to be held in Budapest in June (see <http://helix.nature.com/wcs>). Delegates also

called for the setting up of an international mechanism to address ethical issues raised by scientific developments.

Smile on the face of a distant planet



[WASHINGTON] The Mars Global Surveyor returned this picture of a Martian 'happy face' — actually a 215-kilometre-wide crater called Galle — shortly after beginning routine mapping of the planet's surface last week.

The likeness was seen in Viking pictures in the 1970s, but never gained the same following as the more dour 'face' and 'pyramids' of

the Cydonia region. The surveyor will continue mapping for a full Martian year (687 days), capturing details as small as 1.5 metres across.

nature

the international weekly journal of science

invites applications for the post of

Chief Editor, Biological Sciences

Do you have what it takes to lead a growing team of talented editors publishing the world's finest biological sciences papers? To take on the most demanding and stimulating editorial role in biological sciences publishing? To be responsible for the highest standards in assessment and review? To represent *Nature* in the biological sciences community — and to ensure that *Nature* serves the interests of that community?

If so, we want to hear from you.

The Chief Editor, Biological Sciences, is a senior appointment, reporting directly to the Editor, backed by a strong commitment to investment and expansion. He or she will play a key part in developing *Nature's* publishing policy. Terms and conditions will be highly competitive, reflecting the critical importance and responsibilities of the role.



The ideal candidate will probably have some editorial experience, but will certainly have intellectual vision and strong leadership qualities. He or she will have significant achievements in biological sciences research; demonstrable interests across a wide range of biological sciences; respect in the scientific community; a capacity for prompt and balanced critical assessment; good communications skills; and imaginative ideas for the future development of *Nature* and their contribution to it.

The job is based in *Nature's* modern London offices but will involve international travel to meetings and laboratories.

Applicants should write in confidence, marked 'Personal', to:

Dr Philip Campbell,
The Editor, *Nature*,
4 Crinan St, London N1 9XW, UK.

They should include a CV, a discussion of relevant experience, and an indication of their specific ambitions for *Nature*. Applications should be sent as soon as possible and no later than 16 April.



Words go missing in cyberspace

Sir — Your Briefing on electronic journals asked “Why do researchers download and print?” rather than reading articles on the screen (*Nature* 397, 195–200; 1999). The author suggests that the explanation lies with the poor resolution of text displays on computer screens. I think the real problem lies much deeper, in the way our brains process information, and will not be susceptible to a simple technological fix.

Reading requires us to pile one level of abstraction on top of another. We must recognize letters from the shapes of marks on the page, and the meanings of words from those letters and from the sounds our minds associate with them, and the information in sentences from the meanings of words and from grammatical rules, and still more abstract concepts from the sentences. One feature of reading that makes this achievement possible is that it is grounded in a physical object, the page or book.

One example of the importance of physical context in reading is our positional memory for text — frequently we can easily remember the location of a piece of information on the page and in the book, even though we’ve forgotten its textual content and context. I suspect that this

physical framework is a key part of our reading ability, especially for difficult material.

When we read online we forfeit this level of perception — all text is ephemeral. It flows like water across the screen, and appears and disappears with the click of a mouse. For serious reading of the scientific literature, an abstract may be all we can absorb from the screen without becoming disoriented. It’s as if the ideas we’re reading can’t be given locations in our minds if they don’t have locations in space.

Rosie Redfield

*Department of Zoology,
University of British Columbia,
Vancouver, British Columbia V6T 1Z4, Canada*

Sir — Your Briefing indicates that one promising model for Internet publishing is to ask authors to pay page (as well as figure and table) charges, so reducing the journal’s costs. We see two serious flaws with this approach.

First, as long as there are journals that do not require page charges, journals that do so will compete at a serious disadvantage for quality manuscripts. Second, the charge per page will jeopardize the ability of an

enormous body of authors — namely those from developing nations — to publish in such journals. This is all the more relevant because the Internet frees publishers to a large extent from the onerous page limit restrictions of the printed medium, greatly reducing costs.

We have opted in *Ciencia al Día*, a new electronic-only journal published in Spanish (and soon to be translated into English), to publish free of page charges and at no cost to the reader (<http://www.ciencia.cl/CienciaAlDia/>). This should help to narrow the South-to-North knowledge gap discussed by Vanderlei Canhos and colleagues (*Nature* 397, 201; 1999). In addition, the authors retain their copyrights and can publish the material elsewhere if they desire.

Jorge Golowasch*, Tania Bedrax-Weiss†, Adrián Palacios‡

(Editors, *Ciencia al Día*)

**Volen Center, Brandeis University, Waltham, Massachusetts 02454, USA*

†Computational Intelligence Research Laboratory, University of Oregon, Eugene, Oregon 97403-1269, USA

‡Departamento de Fisiología, Facultad de Ciencias, Universidad de Valparaíso, Valparaíso, Chile

Whales, science and activism

Sir — I hold no brief for Greenpeace, but those of us who monitor these things know that, far from taking action against the stricken whaling ship, as Dan Goodman asserts, the Greenpeace vessel, which happened to be nearby, offered assistance — which was refused — by both radio and signal flags (*Nature* 397, 290; 1999).

Rather than disregarding the advice of its scientific committee, as Goodman asserts, the International Whaling Commission formally accepted it, but has delayed implementation until the whaling nations agree to precautionary measures including independent international inspection and monitoring. What Goodman may find hard to accept is that scientific advice on possible safe catch quotas was nearly all based on research funded by environmental and animal welfare groups, including Greenpeace!

Some organizations in Japan apparently want to resolve the whaling issue. But the Japanese Institute of Cetacean Research, of which Goodman is a member, has made a mockery of its “scientific whaling programme” by resuming, more than half

way through the summer season, the plan to catch the originally decided number of minke whales, despite the fact that this plan was said to have been drawn up to cover all of the designated “sampling” areas and seasons.

Sidney J. Holt

Podere il Falco, Ponticelli PG, 06060, Italy

Many a slip ‘twixt cup and lip

Sir — I read with fascination Len Fisher’s Commentary on the science and art of biscuit dunking (*Nature* 397, 469; 1999). Regrettably, I have not dunked for some years, as the outstanding biscuit for this enjoyment, the traditional English ginger nut, is rare in my country. Danish butter cookies are unsuitably crumbly. However, as a formerly avid dunker in England, I challenge his recommendation of near-horizontal (NH) dunking.

This technique allows dunking only if you abstain from drinking the liquid in your cup. In fact, the minimum dipping angle will exceed the recommended angle if you have navigated your cup through a crowded cafeteria. No: surely it is a question

of simple timing combined with vertically-held (VH) dunking. The same formula applies, but the dunk time must be reduced to one quarter of that of an NH biscuit. To prove this, consider that the desired result is a biscuit stabilized by a dry layer, and that the biscuit is now soaked on both sides.

The desired thickness of remaining dry biscuit (whether lateral or central), however, is also related to the transfer time to the oral cavity, as capillary forces will continue to draw liquid from the supersaturated outer layers into the dry layer after withdrawal from the cup. Even a properly dunked VH biscuit, which is considerably more stable than an NH biscuit (there must be a formula for this), will produce an undesirable mush if the dunking formula ignores cup-to-mouth transfer time.

The problem of ‘asymmetrical’ glazed biscuits, which have an upper surface notoriously less permeable than the lower, needs further research. The physics gets really complicated with the abhorrent dunking practice observed in Latin countries — applying butter and jam to a croissant and then down into the coffee.

Jens C. Jensenius

Department of Medical Microbiology and Immunology, University of Aarhus, DK 8000, Aarhus, Denmark

Lessons from Iraq on bioweapons

There are strong political pressures to relax the scrutiny of suspected biological weapons activity in Iraq. But the experience of United Nations inspectors in the country points to significant dangers in such a policy.

Christian Seelos

The United Nations Special Commission (UNSCOM) was created by the Security Council in 1991 to ensure the destruction or inactivation of all items related to Iraq's biological, chemical and long-range missile weapons capabilities. It was envisaged that Iraq would provide within 15 days a full declaration of all its weapons of mass destruction. Verification of this declaration and disarmament would follow soon after.

But Iraq has fallen well short of this ideal. When UNSCOM's inspectors left Iraq last December in anticipation of military action, the commission's disarmament work was far from complete. Nevertheless, some UN member states are proposing to forgo further disarmament, in favour of establishing a routine mechanism to monitor Iraq's compliance with Security Council resolutions.

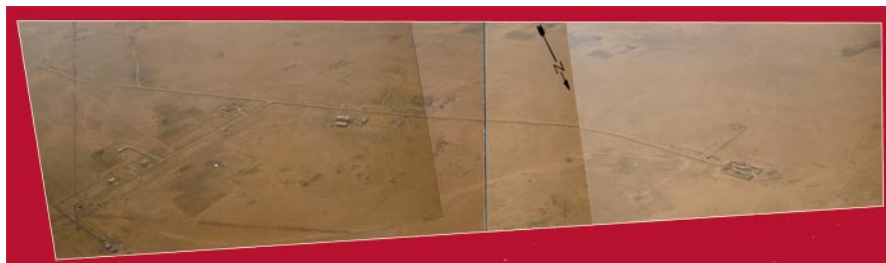
It is part of UNSCOM's mission to develop a plan for future monitoring, but this cannot be effective unless it starts from a baseline of full disclosure of past and present activities. Focusing on Iraq's biological weapons programme, I will argue that UNSCOM's still-fragmentary knowledge makes it essential that the commission be allowed to finish its disarmament work before a monitoring regime can be fully implemented.

Bioweapons factory revealed

In April 1991, in response to its obligations under Security Council resolution 687, Iraq declared that it did not have a biological warfare programme. It also ratified the international convention that prohibits the production of biological weapons. (Iraq would later declare that at that time it had 157 aerial bombs and 25 warheads filled with *Clostridium botulinum* toxin, *Bacillus anthracis* spores and aflatoxin, as well as several thousand litres of biological warfare agents.)

The commission first sent biological warfare experts to Iraq in August 1991. Iraq then declared that it had indeed conducted biological research for military purposes on a limited scale. It claimed these activities had ceased in 1990, that no biological warfare agents were produced in bulk or put in weapons, and that all material had been destroyed.

UNSCOM experts were suspicious that Iraq was not revealing the true scale of its biological warfare programme, but their initial investigations made little progress. The organization had to evolve a more searching mode of operation. Largely because of this approach, and contrary to critics' claims, UNSCOM made significant progress over the



Large-scale preparations for biological warfare: Iraq's main biological weapons site at Al-Hakam (top) housed equipment to produce biological agents (left) that were also placed in aerial bombs.

following years. This led to surprising discoveries about the extent of Iraq's activities.

UNSCOM's knowledge about the programme was still limited to Iraq's initial declaration when the Al-Hakam factory was first inspected in September 1991. The experts were astonished by the factory's size, layout and security measures. It was located in desert 60 kilometres south of Baghdad. The site spread over an area of 3×6 km, secured by a high fence and guard towers. Buildings were widely dispersed across the site.

Iraq claimed that Al-Hakam was a civilian project, producing mainly single-cell protein, based on yeast strains, as animal feed and bio-pesticide (*B. thuringiensis*). But UNSCOM experts had strong suspicions. Their doubts were raised by the isolated location of the factory; the equipment present; the isolation of buildings to create areas of containment; and the dimensions of the animal facility and incinerator. Iraq was also unable to prove the economic viability of the stated projects. UNSCOM concluded that Al-Hakam would have great potential for an offensive biological warfare programme, but there was only circumstantial evidence.

Although Iraq continued to deny that it had ever had an offensive biological warfare programme, the concerns raised by this and other inspections motivated UNSCOM to continue its investigations. Through intrusive inspections, more was learned about Iraq's biological capabilities, and preliminary baselines for future compliance monitoring were set. But these baselines lacked the most important ingredient: a full understanding of the strategy and current status of Iraq's biological warfare programme.

In 1995, the commission became aware that Iraq had imported more than 40 tonnes of bacterial growth media in the late 1980s. That volume was far more than could possibly be justified by industrial activities: Iraq was unable to account credibly for about 17 tonnes. Following continued pressure from UNSCOM, Iraq finally admitted in July 1995 that it had had an offensive biological warfare programme, and had used the missing growth media to produce biological warfare agents. It also admitted that Al-Hakam was a biological warfare factory. But, incredibly, Iraq denied having put biological agents into weapons.

Iraq threatened to cease cooperation with UNSCOM unless there was progress towards the lifting of sanctions. Then, in August 1995, following the departure of the head of Iraq's programmes for weapons of mass destruction, Iraq admitted to a significant biological warfare programme, including research into a range of agents, some of which had been placed in weapons.

After many interviews and site inspections, and a series of "full, final and complete disclosures" by Iraq between 1992 and 1997, UNSCOM gained a better understanding of the programme. In the early 1970s, a biological warfare institute had been established near Baghdad. Although this attempt is alleged to have failed, Iraq says a second biological warfare programme started in 1985, initially under the umbrella of the chemical weapons programme. Shortly afterwards, the group moved to a peninsula southeast of Baghdad, where it operated inside the forensic department of an organization related to Iraq's security apparatus. Small-scale production of agents and toxicity tests with ani-

imals were performed. Research focused on agents including *B. anthracis*, *C. botulinum*, *C. perfringens*, aflatoxins, *Tilletia* spp., trichothecenes, ricin, enteroviruses, rotaviruses and camelpox virus. In addition, as simulants for *B. anthracis*, *B. subtilis* was used in weapon field trials and *B. thuringiensis* was produced for studying spray drying of agents.

In 1988–89 the group moved to the top-secret Al-Hakam factory. In 1990, shortly before the Gulf War, the programme was on the verge of further significant expansion. Research into viruses and genetic engineering began. *C. botulinum*, *B. anthracis* and aflatoxin were produced in quantity and placed in bombs and long-range missile warheads. Devices for large-scale aerosol dissemination of agents were being field-tested in 1991, five months after Iraq had invaded Kuwait.

Destroying the evidence

UNSCOM is the first organization to have demonstrated, through systematic technical work, the existence of a covert biological warfare programme. In 1991, Iraq decided to hide all evidence of its programme, and claims that this was achieved by destruction of all weapons, agents and documents. It continues to argue that all disarmament is therefore finished. But the claimed destruction has not been supported with credible evidence.

The only direct evidence for the programme has been found through UNSCOM inspections. Intact aerial bombs and remnants of missile warheads have recently been unearthed at declared destruction sites. DNA analysis revealed a biological agent. The data only partly fit Iraq's account.

Other evidence was found at the Al-Hakam factory, which was destroyed under UNSCOM supervision in 1996. Iraq later admitted that it had cleaned all the equipment to prevent detection of agents. UNSCOM's sampling of the equipment demonstrated the success of these concealment attempts: no live agent could be isolated. Nevertheless, highly sensitive and specific DNA-based detection methods found traces of agent in some samples, including a pH probe, whose contents escaped decontamination. If done imaginatively and thoroughly, biological sampling is an important tool for inspection teams.

UNSCOM has established that Iraq has the capability to sustain or rejuvenate a significant biological warfare programme. Considering the lengths to which Iraq has gone to limit investigations, it is prudent to assume it still has the intention of acquiring such weapons. So, Iraq is not complying with the technical criteria UNSCOM has set in accordance with Security Council resolutions. The assessment of non-compliance was reached unanimously by experts provided to UNSCOM by many UN member states.

Because this assessment is purely technical and not political, it is not negotiable. It cannot be made to go away by changing the



Desert storm: the Al-Hakam biological weapons factory was destroyed by UNSCOM in 1996.

composition of UNSCOM or by turning it into a new organization — as was proposed in January by several governments, to resolve the stalemate between Iraq and the UN. Iraq's position is that it will not cooperate further with UNSCOM's disarmament work. It demands the lifting of sanctions, claiming that disarmament is finished.

The proposals to change UNSCOM lack technical detail, but are mainly based on the assumption that the preventative benefits of biological monitoring would justify leaving the disarmament work unfinished. On 30 January, the president of the Security Council stated that a new panel would assess all available information on the state of disarmament in Iraq. Recommendations for how to re-establish an effective regime of disarmament and/or monitoring and verification are expected from the panel by 15 April.

Political risk

If we get it wrong in Iraq, what is at stake? Despite strenuous efforts, the commission does not have any credible evidence that Iraq has abandoned its biological warfare programme, destroyed all stocks of agent and weapons, or given up its intention of acquiring weapons of mass destruction. It is unclear whether Iraq's destruction of biological warfare items in 1991 was aimed at obliterating its programme, as claimed, or simply an attempt to conceal the evidence. If the purpose was to deceive the commission, destruction may well have been incomplete. Iraq's intentions are defined by the extent to which it is willing to resolve the 'disarmament' part of the commission's work. Full disclosure, verified by UNSCOM inspectors, is the only credible way for Iraq to prove it has given up its intention of acquiring weapons of mass destruction. In this way, the obliteration of the biological warfare programme can be established and compliance with Security Council resolutions can be demonstrated.

Routine monitoring must be considered primarily as a means of verifying continuing compliance. Because Iraq is violating its obligations under Security Council resolutions, the commission cannot be expected to prove compliance through its biological monitoring, but only to demonstrate the continuation of non-compliance. In the past year alone, inspectors have found 800 signif-

icant pieces of dual-use biological equipment that had not been declared by Iraq.

Several countries have proposed that the commission's unfinished disarmament work (the resolution of which is linked to the lifting of the oil embargo) could be replaced by a comprehensive monitoring system. The implication is that biological monitoring could validate the absence of non-compliance, even though a country is assumed to be violating treaty obligations. Such a concept is not based on any evidence and ignores the relationship between monitoring and compliance. If there is compliance, the absence of non-compliance can be verified with some confidence. If there is no basic compliance, this cannot be done because it is impossible to monitor a whole country continuously.

Biological agent containers and dissemination devices are small and hard to recognize. Equipment used for biological warfare purposes is the same as that used for legitimate academic and industrial activities. Frequent monitoring may be a deterrent, but this is true only for the inspected facility itself.

If the commission's disarmament work is left unfinished this will militate against the establishment of an effective monitoring regime. If Iraq is not forced to make a full disclosure of its past programme, why should it be open about current and future activities? Monitoring against a wall of concealment would constantly trigger alarms indicating potential violations. Keeping the threshold for alarms at a reasonably sensitive level would prevent the building of confidence in the effectiveness of the monitoring system. The next serious crisis between Iraq and the international community is already programmed into such a concept.

Bearing in mind the limitations of monitoring and the necessity to establish compliance, a mature monitoring system can be set up that is not based on unrealistic expectations. It would be based on verification by UNSCOM of a full declaration by Iraq, allowing for a reasonable baseline for future monitoring to be set. Such a system might even be capable of gathering enough evidence to raise the alarm in the case of a violation, although it would probably not stop a country from building a clandestine programme.

If the political will existed to establish a capable monitoring system, it would send a strong message that the international community considers biological weapons to be no part of the arsenal of civilized countries. It would reflect the uncompromising stance that no one will be allowed to get away with acquiring biological weapons. Much more is at stake than just the case of Iraq. □

Christian Seelos is at the United Nations Special Commission, United Nations, S-3027, New York, NY 10017, USA. He is on secondment from the Institute for Tumorbiology—Cancer Research, University of Vienna, Austria. This article represents his personal viewpoint.

Bell's inequality test: more ideal than ever

Alain Aspect

The experimental violation of Bell's inequalities confirms that a pair of entangled photons separated by hundreds of metres must be considered a single non-separable object – it is impossible to assign local physical reality to each photon.

Bell's theorem¹, formulated in 1964, is one of the profound scientific discoveries of the century. Based on the Einstein, Podolsky and Rosen (EPR) *gedanken*, or thought, experiment², it shifted the arguments about the physical reality of quantum systems from the realm of philosophy to the domain of experimental physics. For almost three decades, experimental tests³ of Bell's inequalities have evolved closer and closer to the ideal EPR scheme. An experiment at the University of Innsbruck⁴ has, for the first time, fully enforced Bell's requirement for strict relativistic separation between measurements.

It all started when Einstein *et al.* pointed out that for certain quantum states (described almost simultaneously by Schrödinger, who coined the expression 'quantum entanglement'), quantum mechanics predicts a strong correlation between distant measurements. Figure 1 shows a modern version of the EPR situation, where a pair of entangled photons v_1 and v_2 are travelling in opposite directions away from a source. Results of polarization measurements with both polarizers aligned are 100% correlated. That is, each photon may be found randomly either in channel + or – of the corresponding polarizer, but when photon v_1 is found positively polarized, then its twin companion v_2 is also found positively polarized. Because no signal can connect the two measurements if it travels at a velocity less than or equal to the speed of light, c , and because the choice of the direction of analysis can be made at the very last moment before measurement while the photons are in flight, how — argued Einstein — could one avoid the conclusion that each photon is carrying a property, determining the polarization outcome for any direction of analysis?

This seemingly logical conclusion provides a simple image to understand the correlations between distant and simultaneous measurements. But it means specifying supplementary properties ('elements of reality' in the words of Einstein) beyond the quan-

tum-mechanical description. To the question "Can a quantum-mechanical description of physical reality be considered complete?"² Einstein's answer was clearly negative, but this conclusion was incompatible with the 'Copenhagen interpretation' defended by Bohr, for whom the quantum-mechanical description was the ultimate one⁵. This debate between Einstein and Bohr lasted until the end of their lives. As it was, it could hardly be settled, because there was no apparent disagreement on the correlations predicted for an EPR *gedanken* experiment. The point under discussion was the worldview implied by the analysis of the situation.

Bell's theorem changed the nature of the debate. In a simple and illuminating paper¹, Bell proved that Einstein's point of view (local realism) leads to algebraic predictions (the celebrated Bell's inequality) that are contradicted by the quantum-mechanical predictions for an EPR *gedanken* experiment involving several polarizer orientations. The issue was no longer a matter of taste, or epistemological position: it was a quantitative question that could be answered experimentally, at least in principle.

Prompted by the Clauser–Horne–

Shimony–Holt paper⁶ that framed Bell's inequalities in a way better suited to real experiments, a first series of tests⁷, using photon pairs produced in atomic radiative cascades, was performed in the early 1970s at Berkeley, Harvard and Texas A&M. Most results agreed with quantum mechanics, but the schemes used were far from ideal; in particular, the use of single-channel polarizers only gave access to the + outcome. Progress in laser physics and modern optics led to a new generation of experiments carried out by colleagues and myself at Orsay in the early 1980s. They were based on a highly efficient source of pairs of correlated photons, produced by non-linear laser excitations of an atomic radiative cascade. An experiment involving two-channel polarizers, as in the ideal EPR *gedanken* experiment, gave an unambiguous violation of Bell's inequalities by tens of standard deviations, and an impressive agreement with quantum mechanics⁸.

A third generation of tests, begun in the late 1980s at Maryland and Rochester^{9,10}, used nonlinear splitting of ultraviolet photons to produce pairs of correlated EPR photons. With such pairs, measurements can bear either on discrete variables such as polarization or spin components, as considered by Bell, or on continuous variables of the type originally considered by Einstein, Podolsky and Rosen, and studied at Caltech¹¹. A remarkable feature of such photon sources is the production of two narrow beams of correlated photons that can be fed into two optical fibres, allowing for tests with great distances between the source and the measuring apparatus, as demonstrated over four kilometres in Malvern¹² and over tens of kilometres in Geneva¹³.

The experimenters at Innsbruck⁴ used this method to address a fundamental point raised by Bell. In the experiment shown in Fig. 1, where the polarizers' orientations are kept fixed during a run, it is possible to rec-

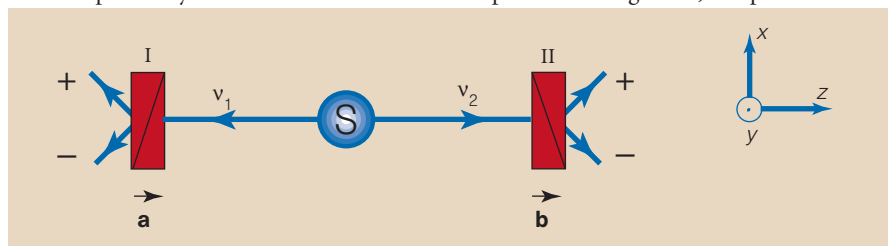


Figure 1 Einstein–Podolsky–Rosen *gedanken* experiment with photons. The two photons, v_1 and v_2 , are analysed by the linear polarizers I and II, which make polarization measurements along $\vec{a}$ and $\vec{b}$ perpendicular to the z axis. Each measurement has two possible outcomes, + or –, and one can measure the probabilities of single or joint measurements at various orientations $\vec{a}$ and $\vec{b}$. For an entangled EPR state, violation of a Bell's inequality indicates that the strong correlations between the measurements on the two opposite sides cannot be explained by an image 'à la Einstein' involving properties carried along by each photon. In the Innsbruck experiment⁴, any possibility of communication between the polarizers, at a velocity less than or equal to that of light, is precluded by random and ultrafast switching of the orientations of the polarizers, separated by a distance of 400 m. On each side, a local computer registers the polarizer orientation and the result of each measurement, with the timing monitored by an atomic clock. Data are gathered and compared for correlation measurements after the end of a run.

oncle the quantum mechanical predictions and Einstein's conceptions by invoking a possible exchange of signals between the polarizers. To avoid this loophole, Bell stressed the importance of experiments "in which the settings are changed during the flight of the particles"¹, so that any direct signal exchange between polarizers would be impossible, provided that the choice of orientations is made randomly in a time shorter than the flight time of the particle or photon, to ensure that relativistic separation is enforced.

Prompted by Bell's remark, a first step towards the realization of this ideal scheme¹⁴ found a violation of Bell's inequality with rapidly switched polarizers, but the polarizer separation (12 m) was too small to allow for a truly random resetting of the polarizers. With a separation of 400 m between their measuring stations, the physicists of Innsbruck⁴ have 1.3 μ s to make random settings of the polarizer and to register the result of the measurement, as well as its exact timing monitored by a local rubidium atomic clock. It is only at the end of the run that the experimentalists gather the two series of data obtained on each side, and look for correlations. The results, in excellent agreement with the quantum mechanical predictions, show an unquestionable violation of Bell's inequalities⁴.

This experiment is remarkably close to the ideal *gedanken* experiment, used to discuss the implications of Bell's theorem. Note that there remains another loophole, due to the limited efficiency of the detectors, but this can be closed by a technological advance that seems plausible in the foreseeable future, and so does not correspond to a radical change in the scheme of the experiment.

Although such an experiment is highly desirable, we can assume for the sake of argument that the present results will remain unchanged with high-efficiency detectors.

The violation of Bell's inequality, with strict relativistic separation between the chosen measurements, means that it is impossible to maintain the image 'à la Einstein' where correlations are explained by common properties determined at the common source and subsequently carried along by each photon. We must conclude that an entangled EPR photon pair is a non-separable object; that is, it is impossible to assign individual local properties (local physical reality) to each photon. In some sense, both photons keep in contact through space and time.

It is worth emphasizing that non-separability, which is at the roots of quantum teleportation¹⁵, does not imply the possibility of practical faster-than-light communication. An observer sitting behind a polarizer only sees an apparently random series of — and + results, and single measurements on his side cannot make him aware that the distant operator has suddenly changed the orientation of his polarizer. Should we then conclude that there is nothing remarkable in this experiment? To convince the reader of the contrary, I suggest we take the point of view of an external observer, who collects the data from the two distant stations at the end of the experiment, and compares the two series of results. This is what the Innsbruck team has done. Looking at the data a posteriori, they found that the correlation immediately changed as soon as one of the polarizers was switched, without any delay allowing for signal propagation: this reflects quantum non-separability.

Whether non-separability of EPR pairs is a real problem or not is a difficult question to settle. As Richard Feynman once said¹⁶: "It has not yet become obvious to me that there is no real problem ... I have entertained myself always by squeezing the difficulty of quantum mechanics into a smaller and smaller place, so as to get more and more worried about this particular item. It seems almost ridiculous that you can squeeze it to a numerical question that one thing is bigger than another. But there you are — it is bigger..." Yes, it is bigger by 30 standard deviations. □

Alain Aspect is in the Laboratoire Charles Fabry, Unité Mixte de Recherche Associée au CNRS, Institut d'Optique Théorique et Appliquée, BP 147-F91405, Orsay, France.
e-mail: alain.aspect@iota.u-psud.fr

1. Bell, J. S. in *Speakable and Unsayable in Quantum Mechanics* 14–21 (Cambridge Univ. Press, 1987).
2. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47**, 777–780 (1935).
3. Aspect, A. in *Waves, Information and Foundation of Physics* Conf. Proc. Vol. 60 (Italian Phys. Soc., Bologna, 1998).
4. Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. *Phys. Rev. Lett.* **81**, 5039–5043 (1998).
5. Bohr, N. *Phys. Rev.* **48**, 696–702 (1933).
6. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. *Phys. Rev. Lett.* **23**, 880–884 (1969).
7. Clauser, J. F. & Shimony, A. *Rep. Prog. Phys.* **41**, 1881–1927 (1978).
8. Aspect, A., Grangier, P. & Roger, G. *Phys. Rev. Lett.* **49**, 91–94 (1982).
9. Shih, Y. H. & Alley, C. O. *Phys. Rev. Lett.* **61**, 2921–2924 (1988).
10. Ou, Z. Y. & Mandel, L. *Phys. Rev. Lett.* **61**, 50–53 (1988).
11. Ou, Z. Y., Peirera, S. F., Kimble, H. J. & Peng, K. C. *Phys. Rev. Lett.* **68**, 3663–3666 (1992).
12. Tapster, P. R., Rarity, J. G. & Owens, P. C. M. *Phys. Rev. Lett.* **73**, 1923–1926 (1994).
13. Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. *Phys. Rev. Lett.* **81**, 3563–3566 (1998).
14. Aspect, A., Dalibard, J. & Roger, G. *Phys. Rev. Lett.* **49**, 1804–1807 (1982).
15. Bennet, C. H. *et al. Phys. Rev. Lett.* **70**, 1895–1898 (1993).
16. Feynman, R. P. *Int. J. Theor. Phys.* **21**, 467–488 (1982).

Conservation

Oryx go back to the brink

A flagship conservation programme, the Arabian Oryx Project in Oman, has suffered a severe setback because of an illegal trade in live animals sold into private collections. The sad story was recounted by Andrew Spalton, a biologist with the project, at a conference in Abu Dhabi earlier this month.

In the early 1960s the Arabian oryx (*Oryx leucoryx*, pictured here) was being hunted to extinction, so a small number were captured to establish breeding herds in the United States and Arabia. The last wild animals were killed in the deserts of Oman in 1972. Ten years later, reintroductions began with the release of ten founder members into Oman's central desert just 75 km from where the last wild oryx had been shot. The liberated oryx flourished, despite serious drought, and by October 1995 there were around 280

animals in the wild, ranging over 16,000 km² of desert.

A few months later the spectre of poaching returned and oryx began to be taken for sale as live animals outside Oman. Nonetheless, the number of



animals continued to increase, to 400 or so, until increasing poaching pressure through 1997 and into 1998 led to a population crash to just 138 in September of last year. At that point the wild population was considered to be no longer viable and 40 animals were taken back into captivity. After further poaching in January of this year, just 11 females and an estimated 85 males remain in the wild.

There is a further reintroduction programme in Saudi Arabia, where poaching is currently less of a threat. So the outlook for oryx in the wild is not entirely grim. But in Oman the situation is bleak, and political action will be needed to remedy matters.

Martyn Gorman

Martyn Gorman is in the Department of Zoology, University of Aberdeen, Aberdeen AB24 3TZ, UK.
e-mail: m.gorman@abdn.ac.uk

MARTYN GORMAN

Telomeres

Ageing hard or hardly ageing?

David Kipling and Richard G. A. Faragher

The potential rewards of understanding ageing are vast, yet the field remains greatly underexplored. Physiological changes in old age — ranging from poor wound healing and a compromised immune system, to cardiovascular problems — can reduce the quality of life for an ever-increasing fraction of the world's population. Reporting in *Cell*, Rudolph *et al.*¹ now provide the first direct evidence that at least some of these changes can be caused by telomere erosion.

Why and how do we age? Ageing is a side effect of optimizing an organism's evolutionary fitness². Because it affects so many different systems, however, it would be naive to expect a single cause or a central 'clock'. The main factors in cognitive decline, for example, may make little contribution to cardiovascular degeneration. Many causes of ageing have been suggested, such as free-radical damage³, replicative senescence⁴ and mitochondrial dysfunction. But these complement rather than exclude each other, and data that support one in no way detract from the potential effects of the others.

One hypothesis concerning ageing of regenerative tissues (as opposed to those that cannot divide) is cell senescence⁴. Normal cells in culture divide a variable, but limited, number of times. They then enter senescence, a state in which they are alive but no longer dividing. Several types of human cell, notably fibroblasts, use the DNA sequences that protect the ends of their chromosomes (telomeres) to count how many divisions they have been through. They can do this because every time a cell goes through DNA replication, a small portion of the end of each telomere is not duplicated, so the telomeres get progressively shorter⁵. This shortening is used as a clock. In experiments where the telomere-maintenance enzyme, telomerase,

which is normally repressed in fibroblasts, was expressed, both telomere shortening and senescence were prevented⁵. Such telomerase-positive cells divide indefinitely in culture, but show no other apparent change in behaviour^{6,7}. The simplest interpretation is that, in fibroblasts at least, human cell senescence is 'telomere driven'.

How could senescent cells contribute to ageing? They have an altered behaviour and pattern of gene expression⁴ from dividing cells, and have been shown to accumulate with age. They may alter their local tissue microenvironment and contribute to the shift from young to aged tissue. But even if they behaved normally, or were rapidly cleared by programmed death, a finite limit to division could still alter tissue function in later life. *In vivo*, cells divide either to maintain routine tissue function or in response to damage or infection. Division rates are species specific and depend on a cell's position within the tissue⁸. Cell division can be frequent, as in the epidermis or intestinal lining, or relatively rare, as with endothelial cells, T cells and fibroblasts, which may divide just once a month or even only once every few years. In such tissues, can a finite limit to division ever be of physiological significance?

Rudolph *et al.*¹ now show that if they reduce the capacity of mouse cells to divide, the function of division-competent tissues is compromised in later life. They followed a cohort of transgenic mice bearing a germline knockout of the *mTR* gene, which encodes an essential component of telomerase. There are few data to suggest that this mutation affects rodent cell senescence, which seems to be triggered through a telomere-independent mechanism⁹. But the continual loss of telomeric sequence does eventually reduce the production of viable cells, probably because,



100 YEARS AGO

Aitken (*Proc. R. S. E.*, vol. ii p. 472, 1882) has given a complete theory of the colour of sea water as observed at various places, based upon the principle that sea water is a blue liquid. According to this view, the green tint often observed in sea water, especially near land, is to be explained by the presence of fine yellow particles. During a recent voyage by the Messageries steamer *Polynésien*, I was permitted, through the kindness of the Commandant Bullard, to erect a tube 736 cm. long against the rail of the after-deck, and to pass through it a continuous stream of water from the ship's salt water service. ... it is useless to comment on most of the results obtained, except in so far as they give a means of easily reproducing the exact tint of pure sea water as seen through a column 736 cm. long. Make up the following solution:- Water, 500 c.c.; Soluble prussian blue, 001 gram; Saturated lime-water just precipitated by the smallest excess of bicarbonate of soda, 5 c.c.. This mixture, when viewed through a tube 18 cm. long, will show with considerable precision the colour of a sample of water from the Mediterranean, lat. 36° 24' N., long. 7° 51' E. of *Paris*.

From *Nature* 16 March 1899.

50 YEARS AGO

An extensive analysis of the energetic disintegrations produced in the silver and bromine nuclei of nuclear emulsions exposed to cosmic radiation indicates that the emitted particles, consisting of α -particles, protons and neutrons, generally escape from the nucleus after the excitation energy has been statistically shared among the constituent nucleons. The disintegration process is then analogous to the evaporation of molecules from a drop of liquid. Occasionally, however, stars are observed in which highly charged nuclear splinters are ejected from strongly excited nuclei before equilibrium has been attained. From a total of six thousand stars, we have observed two examples in which we can identify such heavy fragments. These events are not so rare as this might imply, since their tracks can only be identified when they have a considerable length in the emulsion. The two stars were obtained in 'sandwich' emulsions exposed on the Jungfrauoch (3,500 m.). From *Nature* 19 March 1949.

Table 1 Ageing syndromes in mouse and man

Symptoms	<i>mTR</i> ^{-/-} mice	Werner's syndrome
Shortened division capacity	+	+
Accelerated cell senescence	-	+
Premature greying	+	+
Poor wound healing	+	+
Increased cancer incidence	+	+
Gut defects	+	?
Infertility	+	+
Shortened lifespan	+	+
Decreased adipose tissue	+	+
Hair loss	+	+
Brain changes	-	-
Osteoporosis	-	+
Diabetes	-	+
Atherosclerosis	-	+
Cataract	-	+

if chromosomal ends are no longer protected by telomeres, genomic instability results^{10,11}.

Mice lacking the *mTR* gene show progressive telomere shortening in successive generations. After three generations (G3) of inbreeding, the mice start life appearing normal, but their telomeres are so short that the effects of their compromised division capacity soon become evident. After 18 months, the G3 mice start to show impaired wound healing — a process that requires substantial cell division — after minor surgery. They also show premature greying and hair loss, probably because they cannot maintain the populations of proliferative cells in their hair follicles. By the sixth generation, *mTR*^{-/-} mice show structural changes in the lining of the gut that are intriguingly similar to those seen in normal aged mice¹². These observations strengthen the argument that, in the gut, ageing is associated with a compromised ability to replace cells.

The premature onset of some features of old age is known as a segmental progeroid syndrome. An example is Werner's syndrome, a human premature-ageing disease⁴. The main biological feature of Werner's syndrome is accelerated cell senescence, and the disease has several features in common with the *mTR*^{-/-} mice. In both cases, tissues that cannot divide are largely unaffected. However, the progeroid symptoms are not identical between the two diseases (Table 1), and it will be interesting to compare the *mTR*^{-/-} mice with those carrying a germline knock-out of the mouse Werner gene¹³.

Is there any way to slow the rate of ageing? The only intervention that seems to work reliably in a range of species is calorie restriction. There is evidence that calorie-restricted mice (the human equivalent would be less than 1,500 calories per day) have fewer senescent cells in their tissues¹⁴. It would be interesting to see whether the onset of progeroid features in the *mTR*^{-/-} mice could be delayed by calorie restriction. Whatever the outcome, *mTR*^{-/-} mice are sure to be a powerful tool for analysing the effects of compromised cell replacement on mammalian ageing. □

David Kipling is in the Department of Pathology, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK.

e-mail: KiplingD@cardiff.ac.uk

Richard G. A. Faragher is in the Department of Pharmacy and Biomolecular Sciences, University of Brighton, Cockcroft Building, Lewes Road, Brighton BN2 4GJ, UK.

1. Rudolph, K. L. *et al.* *Cell* **96**, 701–712 (1999).
2. Austad, S. N. *Why We Age* (Wiley, New York, 1997).
3. Wallace, D. C. & Melov, S. *Nature Genet.* **19**, 105–106 (1998).
4. Faragher, R. G. A. & Kipling, D. *BioEssays* **20**, 985–991 (1998).
5. Reddel, R. R. *BioEssays* **20**, 977–984 (1998).
6. Jiang, X. R. *et al.* *Nature Genet.* **21**, 111–114 (1999).
7. Morales, C. P. *et al.* *Nature Genet.* **21**, 115–118 (1999).
8. Chang, E. & Harley, C. B. *Proc. Natl Acad. Sci. USA* **92**, 11190–11194 (1995).
9. Russo, I. *et al.* *Oncogene* **17**, 3417–3426 (1998).
10. Niida, H. *et al.* *Nature Genet.* **19**, 203–206 (1998).
11. Lee, H. W. *et al.* *Nature* **392**, 569–574 (1998).
12. Martin, K., Kirkwood, T. B. L. & Potten, C. S. *Exp. Cell Res.* **241**, 316–323 (1998).
13. Leibel, M. & Leder, P. *Proc. Natl Acad. Sci. USA* **95**, 13097–13102 (1998).
14. Li, Y., Yan, Q. & Wolf, N. S. *Invest. Ophthalmol. Vis. Sci.* **38**, 100–107 (1997).

High-temperature superconductivity

Research enters a new phase

A. J. Millis

“The only difference between the rich and other people is that the rich have more money”, Molly Colum once remarked to Ernest Hemingway. Is the only difference between cuprate (copper-oxide) superconductors and conventional materials the high value of the superconducting transition temperature T_c , or do they exhibit fundamentally new physics? The work of Corson and collaborators reported on page 221 of this issue¹ opens a new experimental window onto this fascinating issue, by showing that it is possible to measure the short-time phase-coherence properties of high- T_c materials.

The technical achievements of this paper are remarkable. Corson and collaborators have measured the frequency-dependent conductivity, which relates the alternating currents flowing in the superconductor to an applied alternating electric field. They have obtained data in the frequency range of 100–600 GHz, which has hitherto been

difficult to access because it is uncomfortably high for conventional microwave measurements and uncomfortably low for conventional optical experiments. Corson *et al.*¹ have developed the coherent time-domain spectroscopy technique pioneered by the groups of Auston and Grischowsky² into a powerful tool for probing this interesting but inconvenient frequency range. In this technique the conductivity is determined from the way in which an electric field pulse is distorted as it propagates through a material. High quality thin-film samples are also essential, and these have recently become available following the success of Eckstein and Bozovic at using the molecular beam epitaxy techniques, familiar from semiconductor processing, to grow transition metal oxides³.

The binding of electrons into ‘Cooper pairs’ causes superconductivity. This binding opens the familiar ‘Bardeen, Cooper and Schrieffer’ gap in the electron energy spectrum, but the fundamental manifestations of

the superconducting state — namely zero resistance and expulsion of magnetic fields — require in addition that Cooper pairs separated by macroscopic distances have the same quantum-mechanical phase. The strength of this long-range phase coherence is usually described by a ‘phase stiffness’ ρ_s , which measures the rigidity of the superconducting state in the same way that the shear modulus measures the rigidity of a solid. If $\rho_s = 0$ the material is not superconducting. The current carried by Cooper pairs is proportional to the gradient of the Cooper pair phase, and so the phase stiffness may be extracted from a conductivity measurement.

In two-dimensional superconductors one may have Cooper pairing without long-range phase coherence. Above a Kosterlitz–Thouless temperature T_{KT} , related to the phase stiffness by $T_{KT} = 8\rho_s/\pi$, long-range phase coherence is destroyed by the presence of vortices (singularities in the phase field corresponding to circulating current patterns). For temperatures only slightly greater than T_{KT} , the material retains phase coherence on short length and time scales, but as T is further increased no trace of phase coherence remains.

This behaviour was essentially observed by Corson *et al.*¹ (Fig. 1). Although their work will undoubtedly stimulate much interesting

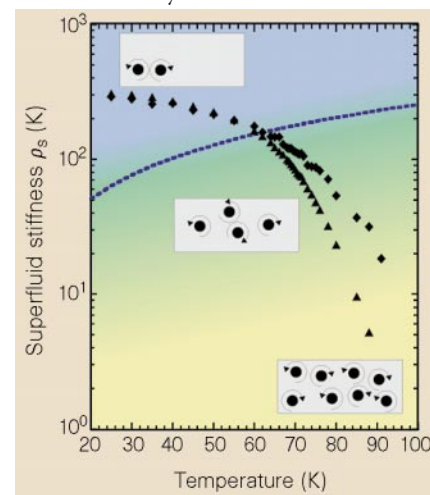


Figure 1 Phase stiffness ρ_s as a function of temperature at frequencies of 100 GHz (triangles) and 600 GHz (diamonds) as determined by Corson *et al.*¹. Dotted line follows the Kosterlitz–Thouless temperature T_{KT} as a function of phase stiffness. The Kosterlitz–Thouless superconducting transition is predicted to occur when ρ_s intersects this line. Insets show behaviour of vortices in the superfluid phase field. At low temperatures, any vortices present are tightly bound in pairs of opposite circulation, so the phase stiffness is large and frequency independent. At $T \approx T_{KT}$ a few unbound vortices are present and ρ_s is non-vanishing at high frequency (short-length scales) but vanishing at low frequency (long-length scales). At $T \gg T_{KT}$, a proliferation of vortices causes ρ_s to vanish at all scales.

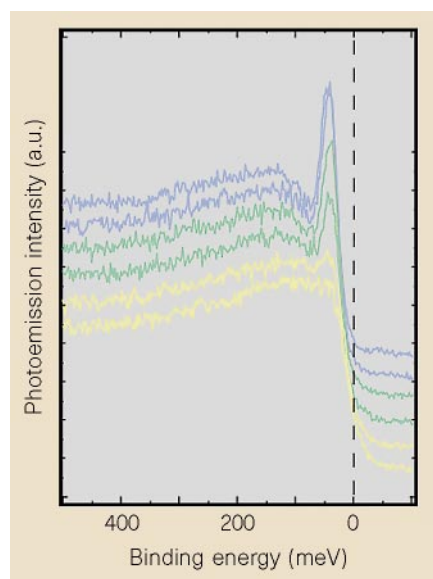


Figure 2 Photoemission spectrum. Photoemission data obtained by Loeser *et al.*⁴, at momentum $p = (\pi, 0)$, as a function of energy at different temperatures for a sample similar to that studied in Fig. 1. Quasiparticle peak (marked by arrow) is evident at $T < T_{KT}$ but not at $T > T_{KT}$. Energy gap, seen as suppression of photoemission intensity in the vicinity of the chemical potential (marked by dotted line at $\omega = 0$) is evident in all data sets, and would cease to be visible at about 150 K.

research into the detailed Kosterlitz–Thouless physics of high- T_c , its fundamental significance is the insight it may give into the new physics of high- T_c superconductors.

The cornerstone of the modern understanding of conventional superconductors is ‘Fermi liquid’ theory, which holds that at the energies and temperatures relevant for everyday life (and for superconductivity), all the complicated physics of interacting electrons can be subsumed into the notion of ‘quasiparticles’: low-energy excitations which behave, for practical purposes, as though they were electrons. Quasiparticles can be observed in angle-resolved photoemission experiments: the data are customarily plotted as a function of energy at fixed momentum and a quasiparticle state shows up as a peak. Figure 2, showing photoemission data⁴ at momentum $p = (0, \pi)$ from a sample similar to the Corson sample shown in Fig. 1, reveals a fundamental mystery. At this momentum, a quasiparticle peak is seen at low temperatures, but as T is increased above T_{KT} the peak vanishes: the superconducting state has quasiparticles, and the non-superconducting state (one hesitates to call it ‘normal’) does not. Figure 2 also shows an energy gap: at this momentum photoemission intensity is suppressed at low energies. The energy gap clearly persists up to higher temperatures than the temperature (~ 100 K) at which all traces of phase coherence are lost, but the presence or absence of

quasiparticles is tied to the phase coherence, not to the gap.

The detailed behaviour of the gap and of the quasiparticle peak depends on the momentum at which the measurement is made, in ways not yet fully elucidated. It is clear, from photoemission and other measurements, that there is a general connection between the onset of apparently unconventional behaviour and the loss of phase coherence. The nature of this connection is one of the pressing issues in high- T_c superconductivity. It is not at all understood theoretically, although it is interesting to note that a relation between phase stiffness and the existence of quasiparticles (and also the possibility of a gap without phase coherence) follows from the ‘resonating valence bond’

picture of Anderson⁵. This picture is not currently in vogue, but perhaps a renewal of interest is indicated. The important point, however, is that the newly available probes of phase coherence and quasiparticle behaviour mean that the issue may soon be clarified by experiment. Exciting times lie ahead. □

A. J. Millis is in the Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA.

e-mail: millis@physics.rutgers.edu

1. Corson, J., Mallozzi, R., Orenstein, J., Eckstein, J. N. & Bozovic, I. *Nature* **398**, 221–223 (1999).
2. Auston, D. H. in *Ultrashort Laser Pulses and Applications* (ed. Kaiser, W.) 183–229 (Springer, Heidelberg, 1988).
3. Eckstein, J. N. & Bozovic, I. *Annu. Rev. Mater. Sci.* **25**, 679–709 (1995).
4. Loeser, A. G. *et al. Phys. Rev. B* **56**, 14185–14189 (1997).
5. Anderson, P. W. *Science* **235**, 1196–1198 (1987).

Agriculture

Neolithic genetic engineering

Svante Pääbo

Of all human inventions, none has had a more profound effect on our history — and on the biosphere as a whole — than agriculture. The agricultural revolution, which began in several regions of the world about 10,000 years ago, allowed food to be produced and stored. This, in turn, permitted the development of large communities, leading to centralization of political power and the emergence of complex societies, so paving the way for large-scale warfare, imperialism, industrialization and almost every other aspect of history as we know it. This momentous development relied on the genetic manipulation of only a handful of plants by early farmers. One of these, maize (or corn), which is now the second largest crop worldwide, was domesticated in middle America around 7,500 years ago from teosinte. This grass looks so different from maize that, until genetic studies showed their close relationship, the two were classified in different genera. For several years, John Doebley and collaborators have hunted the genes that were selected during maize domestication, and they report their latest findings on page 236 of this issue¹.

Almost ten years ago, the Doebley group crossed teosinte and maize and showed² that only five genetic regions are responsible for the main morphological differences between the two. One of these regions encodes a gene called *teosinte branched1* (*tb1*)³. The maize variant of *tb1* causes the side branches of teosinte to shorten and carry ears instead of tassels. Doebley and collaborators⁴ cloned *tb1*, and they have now investigated what effects domestication had on variation in this gene.

In principle, one would expect domestication to reduce genetic variation in maize, because early farmers would have selected

only one or a few teosinte variants with desirable properties. But such an effect has not been seen. In fact, maize is more genetically diverse than many wild plants, indicating that many teosinte alleles made it into the maize gene pool — either during domestication or afterwards, through cross-pollination with teosinte. However, *tb1* is the first gene to be studied that determines a trait early Neolithic farmers were probably interested in, and the results are very different. Doebley and colleagues¹ sequenced a large portion of *tb1* from 17 samples of maize (*Zea mays mays*), 12 samples of *Zea mays parviglumis* (the teosinte species that was probably the direct ancestor of maize), and additional samples from another species of teosinte (*Zea mays mexicana*). They found no drastic reduction in variation in the part of the gene that codes for the *tb1* protein. However, in the upstream, non-transcribed part of the gene, maize contains only 3% of the variation found in teosinte.

When the authors tried to estimate how the DNA sequences evolved, by reconstruct-

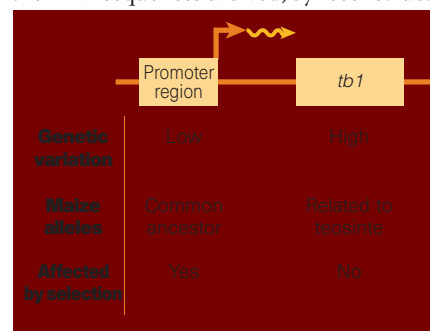


Figure 1 Genetic variation in and around the teosinte branched (*tb1*) gene. Doebley *et al.*¹ have found the greatest variation in the promoter region, which controls the amount of *tb1* expression.



Figure 2 Maize, as we know it today.

ing the most likely relationship among them, the results were drastically different for the upstream and the coding parts of *tb1*. For the coding part, they found that many maize alleles are more closely related to teosinte alleles than to other maize alleles. This is what you would expect if a lot of teosinte variation was incorporated into maize during domestication, as well as subsequently through cross-pollination. But for the upstream part of the gene, the maize alleles trace back to one common ancestor that the maize samples share to the exclusion of almost all teosinte alleles. Furthermore, and equally interesting, the teosinte alleles that are most closely related to the maize alleles are derived from *Z. mays* ssp. *parviglumis*.

So, only the upstream, non-coding part of *tb1* seems to have been affected by strong selection (Fig. 1). This makes sense in view of the gene's effect on maize morphology. The *tb1* protein suppresses growth of lateral branches, and the more of it that accumulates during development, the shorter the branches will be. Thus, a mutation that leads to an increase in the amount of *tb1* produced, rather than a mutation in the coding region, probably controls the change from the teosinte shape to a maize-like morphology (Fig. 2). Because regulatory sequences are generally located upstream of the coding part of the gene, it is reasonable that this would be the part affected by the selection associated with domestication.

But why is the rest of the gene not affected? The answer is recombination, which allowed the selected allele to exchange segments with other alleles in the early maize, and with alleles introduced by cross-pollination from teosinte. In maize, recombination is so frequent that the homogenizing effect of

selection became limited to the upstream region of *tb1*. Moreover, by making several assumptions about the rate of recombination, the frequency of the selected allele in the original teosinte population, and the scale at which the teosinte was grown during domestication, it is possible to gauge the time necessary for selection. These estimates — which obviously depend on the assumptions, and so are tentative — indicate that domestication was rather rapid, happening over some hundreds of years. This is a significant result because archaeologists are still debating how many centuries or millennia were necessary for early farmers to achieve the changes that made maize a mainstay of farming.

Several questions remain. One surprise is that some of the teosinte and maize alleles are almost identical in their upstream regions. So perhaps crucial mutations are located outside the regions that have been studied. Future gene-transfer experiments should clarify that. But this study is fascinating to me because it provides the first glimpse of what went on during one of the earliest genetic-

engineering experiments. It shows that although human activity has left profound traces in the maize gene pool, these changes are restricted to rather small DNA segments surrounding the regions that cause the traits selected by ancient farmers. In future, we will probably be able to find genes that have been targeted for selection by humans in other domesticates, simply by screening their genomes for regions of low variability. By studying such genes, we can learn much about human history. Maize will lead the way, thanks to the work of Doebley and collaborators, but other crops such as wheat, as well as the genomes of domesticated animals such as dogs, cats and cows, are sure to follow. □

Svante Pääbo is at the Max-Planck-Institut für evolutionäre Anthropologie, Inselstrasse 22, D-04103 Leipzig, Germany.
e-mail: paabo@eva.mpg.de

1. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. *Nature* **398**, 236–239 (1999).
2. Doebley, J., Stec, A., Wendel, J. & Edwards, M. *Proc. Natl Acad. Sci. USA* **87**, 9888–9892 (1990).
3. Doebley, J., Stec, A. & Gustus, C. *Genetics* **141**, 333–346 (1995).
4. Doebley, J., Stec, A. & Hubbard, L. *Nature* **386**, 485–488 (1997).

Quantum fluids

Bose gases and their Fermi cousins

Michael R. Andrews

The study of quantum degenerate gases continues to delight physicists, as was evident when researchers met to discuss progress at a workshop last month*. Most of the results concerned Bose–Einstein condensates of dilute alkali gases, a novel type of quantum fluid discovered back in 1995 (refs 1–3) and now the workhorse of what is a rapidly growing field. A new technique for probing a condensate called ‘Bragg spectroscopy’⁴ (W. Ketterle, MIT) is important because it gives us, among other things, the first direct observation of atomic motion within a trapped condensate and yields a quantitative measure of long-range quantum coherence. Another highlight of the meeting was progress reported towards the quantum degeneracy of fermions, notably the cooling of a dilute gas of potassium atoms (D. Jin, JILA). The study of quantum degenerate fermionic systems like this one promises to be just as interesting as, and complementary to, studies of Bose–Einstein condensates, and also to yield new insights into the behaviour of quantum fluids.

Advances in this field (see for example refs 5–8) show that what started out as a sub-

field of laser cooling and atomic physics is now clearly contributing to, and overlapping with, areas such as condensed-matter physics, many-body quantum mechanics, and optics. Further evidence for this overlap, if it were needed, is provided by a study of nonlinear ‘matter-wave’ optics using a Bose–Einstein condensate (Helmerson, NIST) that is reported on page 218 of this issue⁹ (see Box, overleaf).

Prized among a condensate's qualities are its extremely low kinetic energy — measured in mere nanokelvins — and its quantum coherence. But how to measure that energy distribution directly? Until now, *in situ* spatial imaging has been insensitive to the atomic velocities in a trapped condensate, and the ballistic velocities of a released condensate are dominated by the freed interaction energy rather than the initial kinetic energy. The MIT technique of Bragg spectroscopy solves both of these problems. It makes use of a Doppler-sensitive two-photon stimulated Raman technique that coherently ejects atoms from the condensate¹⁰. By precisely varying the frequency difference between the two Raman laser beams, whose wavelengths are 589 nm so as to be nearly resonant with the trapped sodium atoms, one controls which velocity class of condensed atoms is expelled and how fast they leave; this follows rather directly from energy and momentum

*Workshop on Bose–Einstein Condensation and Degenerate Fermi Gases, Center for Theoretical, Atomic, Molecular and Optical Physics, University of Colorado, Boulder, Colorado, USA, 10–12 February, 1999. Talks can be found at <http://condon.colorado.edu/~leland/CTAMOP/February99workshop.html>

conservation. Then, by measuring this outgoing fraction as a function of frequency, one maps out the atoms' original distribution. As testament to the sheer coldness of condensates, Bragg spectroscopy needs sub-kilohertz resolution in the frequency difference between the two lasers. That's because the atoms within move at a mere millimetre or less per second, compared with hundreds of metres per second in a room-temperature gas.

The good news with Bragg spectroscopy is that the momentum spread of atoms in a condensate is limited by the Heisenberg uncertainty relation, that is $\Delta x \Delta p \gtrsim \hbar$. As a result, given the size of the trapped condensate (Δx) one can do nothing more to further reduce the range of momenta (Δp), and in a sense the atoms are maximally coherent; the atoms move with a quantum 'zero point' motion and nothing more. Bragg spectroscopy quantitatively confirms what was previously observed with high-contrast interference patterns in the overlap region of two expanding condensates, which were taken as evidence for long-range quantum

coherence¹¹. The repulsive interatomic interactions, which greatly enlarge the condensate compared with the trap's harmonic oscillator ground state, also have the effect of reducing the zero point motion; this is just Heisenberg's uncertainty relation at work.

Switching gears, one can say that fermions (particles with half-integer units of spin $\hbar$) are the opposites of bosons (particles with integer units of spin $\hbar$), because no two fermions can occupy the same quantum state, whereas bosons like to condense. This quantum-statistical property of fermions, called the Pauli exclusion principle, is responsible for giving us the periodic table of elements, because no two electrons (which are fermions) will share identical orbits, but instead they fill up a shell-like structure in the atom. Now, experimentalists have a quantum degenerate gas of dilute fermions (a stable isotope of potassium; ⁴⁰K) in their toolkit for studying quantum fluids of neutral alkali atoms (D. Jin). Even though fermions don't condense like bosons do, their quantum statistics still yield important consequences. First, in the limit of zero tempera-

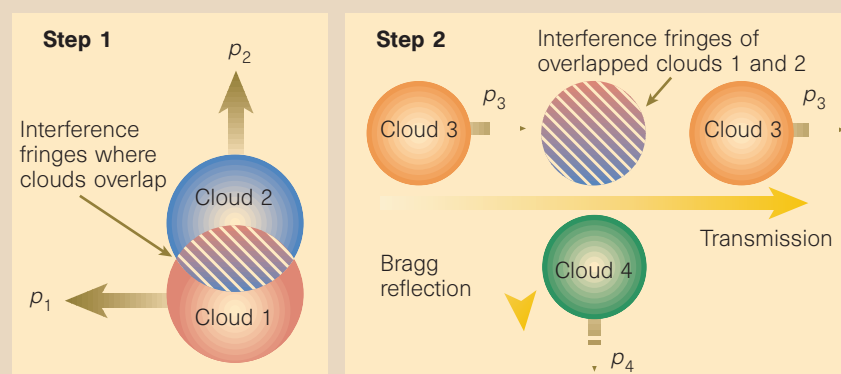
ture, collisions between identical fermions are dramatically suppressed (no *s*-wave, or isotropic, scattering is allowed), which means that evaporative cooling, which relies on collisional thermalization, doesn't work as it does for bosons.

To counteract this suppression of collisions at low temperatures, two different spin states are mixed, allowing collisions between non-identical fermions. This trick maintains thermodynamic equilibrium during evaporation and allows the gas to be cooled all the way down to the microkelvin regime where quantum degeneracy sets in. To be specific, at the workshop the JILA group reported more than 200,000 atoms cooled down to about half a microkelvin in a tight cloverleaf magnetic trap, from which they inferred the sample to be at about half the so-called Fermi temperature — a temperature which represents the energy level that fermions must pile up to even at absolute zero because of Pauli's principle. Although the suppression of collisions was detrimental for cooling, having a gas where the interactions between the particles disappears is quite a novelty — it's the

Nonlinear matter-wave optics

Bose-Einstein condensation made the first atom laser possible, by providing a source of atoms with the required long-range quantum coherence^{10,16}. As described at the Colorado workshop, and by Deng *et al.*⁹ in this issue, researchers at NIST are now taking the field of 'atom optics' a step further and extending nonlinear optics to the 'matter-wave' domain. In analogy to the optical version of the effect, they can exert control over one condensate with another, almost like a switch. The work is important because it opens up a new area in the experimental study of interacting quantum fluids, and implicitly shows how the physics of Bose-Einstein condensates has a multidisciplinary influence.

The optical analogue of the new experiment, called 'four-wave mixing', involves three laser beams overlapping within a nonlinear optical medium: if the three beams satisfy energy and momentum conservation criteria, a fourth laser beam is produced where none was before. The NIST technique involves three overlapping condensates all moving relative to one another so that atoms can be redistributed to form a fourth condensate. Essential to both the optical and atomic versions is the coherence of the waves. With matter waves, however, there is no separate nonlinear medium *per se* — the atoms form their own medium because they interact through collisions. Another



difference is that producing atoms out of thin air is virtually impossible, whereas creating photons is not — creating atoms requires Einstein's rest-mass energy mc^2 , which is essentially infinite on the scale of condensate physics. One exciting area for study opened up by this new work is the possibility of generating non-classical correlations, known as quantum entanglements, between the coherent atomic beams, and this is analogous to similar effects studied with nonlinearly mixed photons.

One can intuitively understand the four-wave mixing of matter waves in the following way: two of the propagating condensates form a quantum interference pattern whose period is given by their relative momentum. The third condensate then scatters from this matter-wave grating,

owing to the nonlinearity of the interactions, and is thus Bragg reflected to form the fourth beam (see figure). Conceptually, it helps to spatially separate the clouds and consider the four-wave mixing process — at least the version presented by Deng *et al.*⁹ in their Fig. 1b on page 219 — as occurring in two somewhat artificial steps. The first step has condensate 1 (with momentum p_1) and condensate 2 (with momentum p_2) producing diagonal interference fringes. In the second step, condensate 3 (with momentum p_3) is Bragg reflected by those fringes to form condensate 4 (with momentum p_4) going off at a right angle. From symmetry considerations, because the atoms are identical bosons the same description holds when exchanging the roles of condensates 1 and 3.

M. R. A.

sort of 'ideal' system physicists love to play with — and was previously proposed as a natural way to get rid of collisional frequency shifts in atomic clocks¹². This is potentially an important application, because these shifts are one of the current limitations.

Predictions for the optical properties of quantum degenerate fermionic gases are intriguing, particularly because they involve an effect previously seen in cavity quantum electrodynamics. There, an excited-state atom could be inhibited from spontaneously emitting a photon if the resonator surrounding it had no compatible electromagnetic modes¹³. With a quantum degenerate fermionic gas, however, it could be the cloud itself that inhibits emission, because the excited atom would have no momentum state available to recoil into; all those states could already be occupied by other atoms in the gas^{14,15}. Measuring the angular and spectral dependencies of scattered light are some of the techniques suggested for gaining access to this and other differences between the optical properties of quantum degenerate Fermi and classical gases.

It is worth noting that at temperatures below the onset of quantum degeneracy,

one would expect the pairing of attractive fermions into a Bardeen–Cooper–Schrieffer superfluid-like state, where the pairs gain a bosonic character and can then form a sort of condensate. But that's an entirely different effect and beyond the scope of current results. □

Michael R. Andrews is at Bell Laboratories, Lucent Technologies, 700 Mountain Avenue, Murray Hill, New Jersey 07974, USA.

e-mail: mikea@bell-labs.com

1. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Science* **269**, 198–201 (1995).
2. Davis, K. B. *et al. Phys. Rev. Lett.* **75**, 3969–3973 (1995).
3. Bradley, C. C., Sackett, C. A., Tollett, J. J. & Hulet, R. G. *Phys. Rev. Lett.* **75**, 1687–1690 (1995).
4. Stenger, J. *et al. http://xxx.lanl.gov/abs/cond-mat/9901109*
5. Anderson, B. P. & Kasevich, M. A. *Science* **282**, 1686–1689 (1998).
6. Stenger, J. *et al. Nature* **396**, 345–348 (1998).
7. Hall, D. S., Matthews, M. R., Wieman, C. E. & Cornell, E. A. *Phys. Rev. Lett.* **81**, 1543–1546 (1998).
8. Hau, L., Harris, S., Dutton, Z. & Behroozi, C. H. *Nature* **397**, 594–598 (1999).
9. Deng, L. *et al. Nature* **398**, 218–220 (1999).
10. Kozuma, M. *et al. Phys. Rev. Lett.* **82**, 871–875 (1999).
11. Andrews, M. R. *et al. Science* **275**, 637–641 (1997).
12. Gibble, K. & Verhaar, B. J. *Phys. Rev. A* **52**, 3370–3373 (1995).
13. Haroche, S. & Kleppner, D. *Phys. Today* **42**, 24–30 (1989).
14. Busch, T., Anglin, J. R., Cirac, J. I. & Zoller, P. *Europhys. Lett.* **44**, 1–6 (1998).
15. DeMarco, B. & Jin, D. S. *Phys. Rev. A* **58**, R4267–R4270 (1998).
16. Mewes, M. O. *et al. Phys. Rev. Lett.* **78**, 582–585 (1997).

Biotechnology

A new-for-old urinary bladder

Jeffrey A. Hubbell

In an astonishing feat of tissue engineering, described in last month's *Nature Biotechnology* by Oberpenning *et al.*¹, a functional dog bladder has been reconstructed from cells that were obtained by biopsy of the bladder wall and guided over degradable polymers built into a bladder shape. This achievement may open the way for bladder replacement in patients who have lost their own bladders as a result of disease².

First, Oberpenning *et al.* isolated smooth-muscle cells from the outside and urothelial cells from the inside of the bladder wall, then expanded them in culture for four weeks before seeding the smooth-muscle cells onto the outer surface of a bladder-shaped polyester textile and the urothelial cells onto the inner surface. Finally, they implanted these constructs into dogs who had had most of their own bladders removed — and the new bladder wall performed successfully for the next year, enabling the reconstructed bladder to substitute for the real thing.

Early work on cell transplantation for the replacement of muscular tissues³ explored the behaviour of smooth-muscle cells in collagen matrices. But the smooth-muscle cell has an annoying habit of transforming itself both in culture and under a variety of conditions *in vivo* from its normal contractile phenotype into a secretory phenotype

known as a myofibroblast⁴, which expresses and secretes large amounts of collagen, leading to the formation of a scar. Rapid healing is good in that it allows an early resumption of function, but bad in that scarring may interfere with optimal ultimate function. The myofibroblast is the enemy that sabotages the attempts of the tissue engineer to reconstruct muscular organs, because this untoward differentiation often results in scar

tissue instead of the desired functional tissue.

In the controls used by Oberpenning *et al.* healing was apparent with disorganized scar formation. When most of the bladder was removed (the trigone area, which comprises the plumbing for inflow and outflow, was always spared) and replaced by a textile implant without cells, a poorly organized, muscle-poor bladder wall formed that was lined with a more-or-less normal urothelium (compare Fig. 1a and b; the urothelium is the darker-staining stripe at the top). Although the urothelium has a high regenerative potential and so can often be extremely helpful to the tissue engineer, here this was inadequate and the resulting bladder walls were far too contracted and stiff to function like normal bladders.

When the wall of the textile implant was seeded with autologous bladder-wall smooth-muscle cells and urothelial cells (Fig. 1c), however, the morphology of the resultant bladder wall was remarkably like that of the normal bladder (Fig. 1a), with clearly visible muscle (dark staining towards the bottom of Fig. 1a, c) and urothelium. This tissue-engineered neo-bladder functioned much like the natural bladder, demonstrating stability in the volume of urine it retained and in its compliance (a measure of the stress–strain behaviour of the bladder wall *in situ*).

Why did the seeded smooth-muscle cells of Oberpenning *et al.* apparently retain a contractile phenotype? The answer is that we do not know. Perhaps there is a dialogue that goes on between the urothelium and the underlying smooth muscle, rather as there is between the endothelium and smooth muscle in blood vessels⁵. This vascular dialogue has also been exploited by tissue engineers, who seeded autologous vascular smooth-muscle cells within a collagen matrix in a vascular graft, then seeded the graft with

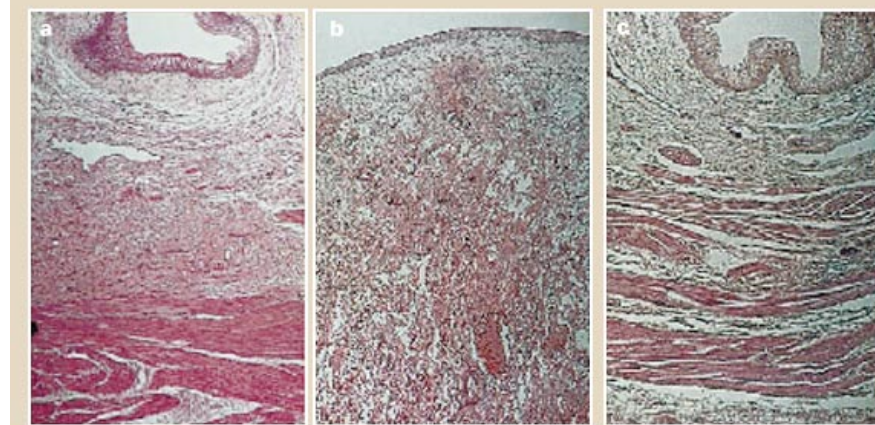


Figure 1 Histology of the urinary bladder implants after surgery. a, Bladder wall in a normal canine; b, after implantation for six months of an initially cell-free degradable polyester bladder scaffold; and c, after implantation for six months of a scaffold originally seeded with smooth-muscle cells and urothelial cells. In a–c, a urothelium is apparent at the top (dark staining; thicker in a and c); in a and c, smooth muscle is visible (dark staining) in the bottom third (a) and half (c) of the images.

autologous endothelial cells, cultured it, and implanted it into dogs⁶.

It was found that when the construct contained both cell types, there was a gradual switch in the smooth-muscle cells from their secretory phenotype (myofibroblasts) to the contractile phenotype, and the vessel wall reached a limited thickness. But when the endothelial cells were omitted from the graft, this did not happen and the vessel wall continued to thicken over time. Some crosstalk between the endothelial cells and the underlying smooth muscle, perhaps involving nitric oxide, was apparently at play in inducing this critical switch, and a similar effect may have been operating in the reconstruction experiments of Oberpenning *et al.*

Where does tissue engineering go from here? The field is admittedly still pre-pubescent. Soon, patient biopsies will be used routinely to isolate stem cells, which presumably have a vastly superior regenerative capacity. The nutritional requirements of thicker tissues for building up muscle or other cells simultaneously with a vascular network will need to be addressed, together with the problem of innervation, perhaps by making use of the available repertoire of growth factors and the genes needed to express them⁷. It is grati-

fying to see these pre-pubescent skills maturing to a point where they should soon have clinical relevance and application.

It might be argued that the observations of Oberpenning *et al.* are merely phenomenological, but we must remember that many fields started off like this — developmental biology, for example, began as a study of phenomenology in mutants and has ended up dissecting molecular mechanisms. Tissue engineering may become similarly refined. And even if the results reported by Oberpenning *et al.* are phenomenological, their implications are still phenomenal: one day, tissue engineers may be able to fashion more complex organs as well. □

Jeffrey A. Hubbell is at the Institute for Biomedical Engineering and Department of Materials, Swiss Federal Institute of Technology (ETH) and University of Zurich, Moussonstrasse 18, CH-8044 Zurich, Switzerland.

e-mail: hubbell@biomed.mat.ethz.ch

1. Oberpenning, F., Meng, J., Yoo, J. J. & Atala, A. *Nature Biotechnol.* **17**, 149–155 (1999).
2. Langer, R. & Vacanti, J. P. *Science* **260**, 920–926 (1993).
3. Weinberg, C. B. & Bell, E. *Science* **231**, 397–400 (1986).
4. Schurch, W., Seemayer, T. A. & Gabbiani, G. *Am. J. Surg. Pathol.* **22**, 141–147 (1998).
5. Mattson, E. J. *et al. J. Vasc. Surg.* **28**, 514–521 (1998).
6. Hirai, J. & Matsuda, T. *Cell Transplant.* **5**, 93–105 (1996).
7. Hubbell, J. A. *Bio/Technology* **13**, 565–576 (1995).

Quantum gravity

Testing time for theories

D. V. Ahluwalia

Theories of quantum gravity attempt to combine quantum mechanics — which reigns supreme at the atomic scale — and the classical theory of general relativity, which governs planets and galaxies. The era of quantum gravity began about 50 years ago with the pioneering work of Chandrasekhar, when it was realized that quantum mechanics and gravitation combine in a fundamental manner to form white dwarfs — the Earth-sized stars of roughly a solar mass. That singular discovery had to wait another 25 years for the excitement to begin afresh with ideas put forward by Bekenstein, Hawking, Unruh and Wheeler, among others. Yet, the present activity runs the danger of evolving into a mathematical science fiction in the absence of an experimental counterpart. For the last quarter of a century an important, although limited, experimental quantum-gravity programme has existed. Giovanni Amelino-Camelia's paper on page 216 of this issue¹ adds to this in a fundamental manner by suggesting that the fuzziness of space–time is experimentally accessible. Using existing data from a certain type of gravity-wave detector he is able to set significant bounds on proposed quantum properties of space–time. Instruments under construction promise to take us further still.

If the Universe is structured the way theorists believe, the quantum mechanical unification of non-gravitational and gravitational interactions must occur at a particular energy scale, called the Planck scale — just as the electromagnetic, strong and weak nuclear forces are predicted to unify at a particular energy scale. But the actual scale at which theories of quantum gravity can be explored has been poorly constrained. The extreme smallness of both the Planck length ($\sim 10^{-35}$ m), on the one side, and the ratio of the gravitational to the electrical forces acting between, say, two electrons, on the other side (2.4×10^{-43}), has led to the widespread belief that the realm of quantum gravity is beyond terrestrial experiments. But the following elementary facts can help dispel this view. For terrestrial experiments, a highly relevant dimensionless quantity is derived from the Newtonian gravitational potential ϕ_{grav} and the speed of light c . The constructed dimensionless object is:

$$\Phi_{\text{grav}} \equiv [\phi_{\text{grav}}/c^2]_{\text{Earth}} = 6.95 \times 10^{-10}$$

This is about 33 orders of magnitude larger than the ratio of the gravitational to the electrical force. Moreover, for quantum-gravity experiments it is not only the force that is of relevance, but also the phases of the wave

functions of quantum mechanical states. Although in the non-relativistic weak-field limit, the force depends on the gradient of ϕ_{grav} , the phase depends directly on ϕ_{grav} . In many situations, phase measurements give an enormous advantage and can also highlight the differences between the quantum and classical realms of gravity. Therefore, for a physical state appearing as a linear superposition of different mass (or energy) eigenstates, non-trivial, gravitationally induced phases can come into play, and in principle are measurable.

Gravity waves are an important prediction of Einstein's theory of general relativity and, although not yet detected directly, may provide information about the very early Universe. Although the Planck scale is extremely small, modern gravity-wave interferometers are designed to detect minute displacements in the positions of some test masses (relative to a beam-splitter). There are additional considerations that may counterbalance the smallness of the Planck length. In certain quantum-gravity theories the quantity $c^2 f^{-2} \lambda_{\text{Planck}}$ (f = frequency) characterizes the fuzziness in the distance between the test masses used in the experiment, whereas the quantity $cS(f)^2$ (where $S(f)$ is the amplitude spectral density of the gravity-wave interferometer) characterizes the level of sensitivity to such fuzziness that modern gravity-wave detectors can achieve. For frequencies of a few hundred hertz these two quantities are comparable.

An experiment that exploited some of these observations, and which still remains relatively unknown in the quantum-gravity community, is the 1975 experiment of Colella, Overhauser, and Werner (COW)². It simultaneously explored the quantum and gravitational realms using neutron interferometry. To the accuracy in phase shifts of about 1% available at that time, the COW experiment established the non-relativistic weak-field limit of any viable theory of quantum gravity. For a single mass eigenstate, the amplitudes associated with each of the paths in a COW interferometer (to be distinguished from interferometers designed to detect gravity waves) pick up a different gravitationally induced phase, and so result in an observable change in the interference pattern as one path is rotated with respect to the other. In this table-top experiment the rotation is carried out in such a manner that each path experiences a different gravitational potential.

However, as the experimental accuracy has improved, the latest (1997) experiments show that a statistically significant discrepancy has begun to appear between the theoretical prediction of the gravity-induced phase shift and that which is experimentally observed³. On the other hand, when one studies the effects of gravity in certain atomic systems, where the experimental measure-

ments are about five orders of magnitude better, no statistically significant discrepancies are seen⁴. In the latest such experiments, gravitationally induced quantum effects of the local tides at Stanford have been observed.

Is there a slight difference in the manner in which neutrons and electrons interact with gravity, and can this point the way towards an appropriate expression of quantum gravity? In recent years, while the new generation of COW experiments was reaching these higher levels of insight, three more quantum-gravity experiments have been proposed. These experiments are based on the possibility that quantum gravity might affect the nature of fundamental symmetries, or that the theory of general relativity itself may not provide a complete description of gravitation. One experiment is based on tests of CPT (the combined symmetries of particle–antiparticle, parity and time reversal) invariance using the very sensitive neutral-kaon system⁵. The second concerns the possibility that quantum gravity might deform Lorentz symmetries in a way that would alter the propagation of the γ -rays we collect from astrophysical sources⁶. The third test explores the incompleteness of the theory of general relativity itself, and proposes to measure quantum mechanically any constant gravitational potential in which we may be embedded⁷.

Here we come to the Amelino-Camelia paper¹, which opens up yet another realm of quantum gravity to experiments. The experiments he proposes would probe what is perhaps the central element of quantum gravity, namely, the concept of space–time. He notes

that almost all existing approaches to quantum gravity expect a modification of the classical picture of space–time by introducing some sort of fuzziness. With that we all agree, but there are various proposals as to what concrete form this fuzziness takes. Amelino-Camelia convincingly argues that gravity-wave interferometers have precisely the right ability to probe the fuzziness of space–time. He sets highly significant limits (in one case ultra-Planckian) on the length scales characterizing two fuzziness situations. Amelino-Camelia's paper exploits the differences in the fuzziness of the distance between the test masses used in the experiment, as predicted by different theories of quantum gravity. It turns out that these differences are large enough to be detected in the experiments being proposed. So, whereas the smallness of the Planck scale and the ratio of gravitational to electrical forces does play its own essential role in nature, it does not make quantum gravity a science where humans cannot venture to probe her secrets. □

D. V. Ahluwalia is at the Escuela de Física, Universidad Autónoma de Zacatecas, Apartado Postal C-580, Zacatecas, ZAC 98068, Mexico. e-mails: ahluwalia@phases.reduaz.mx, av@p25hp.lanl.gov

1. Amelino-Camelia, G. *Nature* **398**, 216–218 (1999).
2. Colella, R., Overhauser, A. W. & Werner, S. A. *Phys. Rev. Lett.* **34**, 1472–1474 (1975).
3. Littrell, K. C., Allman, B. E. & Werner, S. A. *Phys. Rev. A* **56**, 1767–1780 (1997).
4. Young, B., Kasevich, M. & Chu, S. in *Atom Interferometry* 363–406 (Academic, New York, 1997).
5. Mavromatos, N. E. et al. (CPLEAR Collaboration) *Phys. Lett. B* **364**, 239–245 (1995).
6. Amelino-Camelia, G., Ellis, J., Mavromatos, N. E., Nanopoulos, D. V. & Sarkar, S. *Nature* **393**, 763–765 (1998).
7. Ahluwalia, D. V. *Mod. Phys. Lett. A* **13**, 1393–1400 (1998).

Transcriptional regulation

Modification by nuclear export?

Patrick G. Hogan and Anjana Rao

The length of time that transcription factors remain active in the nucleus is critical for regulating gene transcription. Transcription factors — or enzymes that modulate their activity — are transported into or out of the nucleus by proteins that recognize nuclear-localization signals (NLS) and nuclear-export signals (NES), respectively^{1,2}. There are several ways to regulate the relative accessibility of NES and NLS: by post-translational modification of nearby residues; by changes in protein structure; or by interaction of the NLS or NES with specific regulatory proteins. On page 256 of this issue, Zhu and McKeon³ invoke a variation of the third mechanism to explain the regulated nuclear export of NF-AT4, a transcription factor that has both an NLS and an NES.

Since they were identified a few years ago, proteins of the NF-AT family have been attractive models for studying nuclear

import and export^{4–10}. In resting cells, NF-AT proteins reside in the cytoplasm — their NLS is masked, directly or indirectly, by phosphorylation of specific serine residues in the NF-AT regulatory domain^{11,12}. In stimulated cells, an increase in intracellular calcium activates a phosphatase called calcineurin, which dephosphorylates the 'masking' residues, exposing the NLS and initiating nuclear import. In this way, the NLS is exposed and the NF-AT proteins enter the nucleus. But if the activity of calcineurin drops, NF-AT proteins are exported from the nucleus and NF-AT-dependent transcription stops^{5–7}. Nuclear export is accompanied by rephosphorylation of the NF-AT regulatory domain^{5,6}, and it is an active process^{3,9,10} mediated by Crm1, an export receptor that recognizes the leucine-rich class of NES^{1,2}.

Zhu and McKeon³ now show that calcineurin antagonizes the Crm1-mediated

nuclear export of NF-AT4. They used a deleted version of NF-AT4 (NF-AT4 Δ Z), which lacks the NLS-masking sequence and is constitutively found in the nucleus¹². When the authors overexpressed Crm1, NF-AT4 Δ Z was relocated to the cytoplasm, suggesting that, as well as having an exposed NLS, this protein bears an exposed NES. But when they added active calcineurin to the picture, Crm1 could no longer export NF-AT4 Δ Z. The authors conclude that calcineurin regulates the accessibility of both the NF-AT4 NLS and its NES.

In a surprising twist, Zhu and McKeon find that calcineurin antagonizes Crm1-mediated export not by virtue of its phosphatase activity, but rather by competing with Crm1 for binding to NF-AT4 (Fig. 1a). They show that catalytically inactive calcineurin inhibits the binding of NF-AT4 to Crm1 *in vitro*, and that overexpression of this form of calcineurin antagonizes Crm1-mediated export of NF-AT4 Δ Z just as effectively as overexpression of the active enzyme.

The authors further show that whereas NF-AT4 Δ Z requires an intracellular calcium signal to be transcriptionally active, a more extensively deleted NF-AT4 lacking the NES does not. Their interpretation is that Crm1-mediated export, and competition between Crm1 and calcineurin, determine whether or not nuclear NF-AT4 is incorporated into a productive transcriptional complex. This conclusion will be controversial. In fact, rapid Crm1-mediated export cannot, by itself, account for the poor transcriptional activity of NF-AT4 Δ Z. There is plenty of NF-AT4 in the nucleus, and its transcriptional activity should reflect its nuclear concentration — unless a fraction of the NF-AT4 is unavailable for formation of transcriptional complexes.

There are at least two ways to account for the transcriptional data. One way is to extend the idea of competition between Crm1 and calcineurin by suggesting that when NF-AT4 Δ Z is complexed with calcineurin it can bind DNA and stimulate transcription, whereas when complexed with Crm1 it cannot. A second possibility, which is attractive because it does not require stoichiometric binding of calcineurin to NF-AT4 Δ Z, is that the active calcineurin further dephosphorylates NF-AT4 Δ Z. The result could be a conformational change in NF-AT4 Δ Z, which increases its transcriptional activity.

Implicit in Zhu and McKeon's model is the assumption that NF-AT4 Δ Z is a good mimic for dephosphorylated NF-AT4. Indeed, the phosphoserine residues that mask the NLS have been eliminated (albeit by deletion rather than point mutation). But these are not the only residues that could be dephosphorylated by calcineurin. So, although NF-AT4 Δ Z might mimic a partially dephosphorylated form of NF-AT4, it

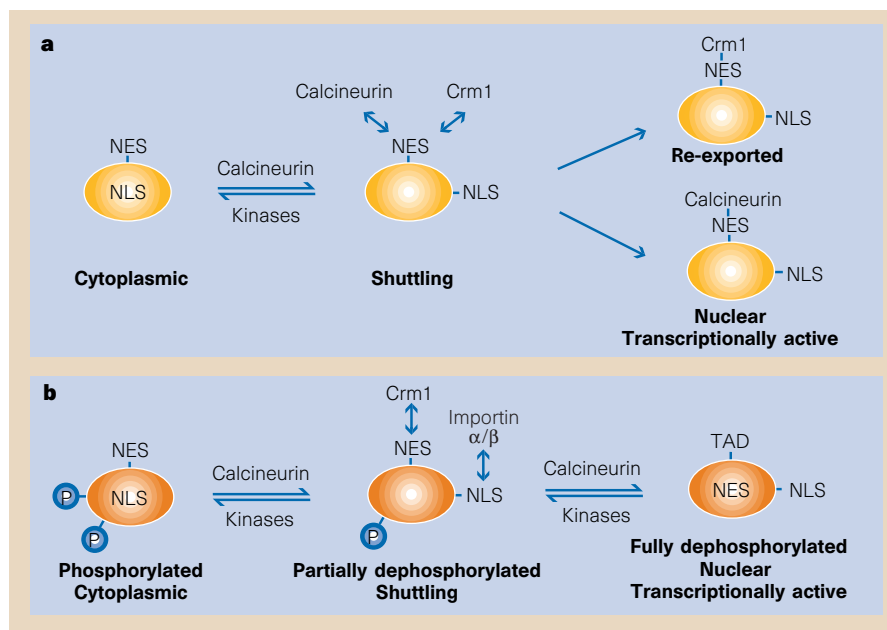


Figure 1 Regulation of the nuclear import and export of NF-AT4. **a**, The model of Zhu and McKeon³. In activated cells, import is initiated by calcineurin-mediated dephosphorylation of NF-AT4. This results in exposure of the nuclear-localization signal (NLS), yielding a protein with an exposed nuclear-export signal (NES) and an exposed NLS. Calcineurin and Crm1 compete for binding to the NES. Crm1-bound NF-AT4 is re-exported from the nucleus. But binding of active calcineurin to NF-AT4 blocks the access of Crm1 to the NES, and the NF-AT4 is incorporated into an active transcriptional complex. **b**, Alternative model. Partial dephosphorylation uncovers the NLS but does not confer full transcriptional activity. Full activation requires further dephosphorylation, which may result in increased affinity for DNA, increased exposure of the transactivation domain (TAD) and, possibly, blocking of the NES.

probably does not resemble the transcriptionally active, fully dephosphorylated form. It would not be surprising if, by further dephosphorylation, calcineurin causes changes that increase the transcriptional capacity of NF-AT4ΔZ (Fig. 1b). There is good evidence that, on full dephosphorylation, another family member (NF-AT1) is converted to a form with a higher affinity for DNA and increased transcriptional function^{4,5,13}.

How generalizable is Zhu and McKeon's model? Can it be applied to other members of the NF-AT family? The NF-AT proteins are all controlled by calcineurin, and the overall organization of the NF-AT regulatory domain is conserved in all four proteins. However, the regulatory domains show considerable primary sequence diversity — in particular, NF-AT2 and NF-AT4 differ in the locations of their NLS-masking sequences^{8,12} and NES^{3,9}, and distinct kinases seem to be involved in their export^{11,12}. In other words, the NF-AT proteins are likely to differ in the details of their regulation, and we must be careful when making generalizations. It is conceivable, however, that the basic characteristic of the model (calcineurin–Crm1 competition) is preserved throughout the family. The calcineurin-binding surface of NF-AT is likely to encompass not only the major calcineurin-binding loop, conserved in all four NF-AT proteins¹⁴, but also regions

that are not contiguous in the primary sequence. Competition between calcineurin and Crm1 will occur if the calcineurin- and Crm1-binding sites overlap, regardless of where the NES is in the primary sequence. It will be a challenge to decipher which regulatory mechanisms are common to all NF-ATs, and which are unique to each family member.

Patrick G. Hogan and Anjana Rao are at the Center for Blood Research and Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.
e-mails: hogan@cbr.med.harvard.edu
arao@cbr.med.harvard.edu

- Ohno, M., Fornerod, M. & Mattaj, J. W. *Cell* **92**, 327–336 (1998).
- Nigg, E. A. *Nature* **386**, 779–787 (1997).
- Zhu, J. & McKeon, F. *Nature* **398**, 256–260 (1999).
- Shaw, K. T.-Y. et al. *Proc. Natl Acad. Sci. USA* **92**, 11205–11209 (1995).
- Loh, C. et al. *J. Biol. Chem.* **271**, 10884–10891 (1996).
- Shibasaki, F., Price, E. R., Milan, D. & McKeon, F. *Nature* **382**, 370–373 (1996).
- Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. *Nature* **383**, 837–840 (1996).
- Beals, C. A., Clipstone, N. A., Ho, S. N. & Crabtree, G. R. *Genes Dev.* **11**, 824–834 (1997).
- Klemm, J. D., Beals, C. A. & Crabtree, G. R. *Curr. Biol.* **7**, 638–644 (1997).
- Kehlenbach, R. H., Dickmanns, A. & Gerace, L. J. *Cell Biol.* **141**, 863–874 (1998).
- Beals, C. A., Sheridan, C. M., Turck, C. W., Gardner, P. & Crabtree, G. R. *Science* **275**, 1930–1933 (1997).
- Zhu, J. et al. *Cell* **93**, 851–861 (1998).
- Luo, C., Burgeon, E. & Rao, A. *J. Exp. Med.* **184**, 141–147 (1996).
- Aramburu, J. et al. *Mol. Cell* **1**, 627–637 (1998).

Daedalus

Sweeping the seas

The polar seas abound in krill, the tiny crustacea that live on plankton. The main predator of krill is the whale; now that the whaling industry has driven whales almost to extinction, the krill is richly abundant. So Daedalus plans to harvest it directly.

What is needed, he says, is some sort of automatic, self-propelled submerged vehicle that would wander at random around the polar seas, sucking in the krill like a giant vacuum-cleaner. When fully laden, it would return with its catch to a home port or a mother-ship. Modern robotics and satellite navigation techniques make this fairly simple in principle. But the devil, as ever, is in the details.

Daedalus's 'oceanic filter-feeder' is essentially an open framework whose bow is a huge filter. Its mesh is chosen to admit krill, but to exclude bigger fish, weed, flotsam and so on, which would complicate the processing downstream. As the craft travels round the ocean, it will strain the krill continuously from its densest stratum just under the surface. The main problem, of course, is that each voyage will continue for many days or weeks; the captured krill will die and rot long before it reaches port. Hence the need for processing: some form of on-board chemical engineering to preserve the catch. Daedalus's design cunningly splits the krill intake into two streams. The smaller stream will indeed be allowed to die and rot. It will ferment to methane gas, the fuel for the propulsion and electronics of the craft, and to carbon dioxide, which will be catalytically reacted with methane to give acetic acid (vinegar). The vinegar will pickle the larger stream of krill, preserving it in a huge flexible plastic 'stomach'. When sensors on the stomach detect that it is full, they will switch the navigation system from 'cruise' to 'homing', and the filter-feeder will return to base with its catch.

Bulk pickled krill will be an intriguing new resource. Since filter-feeders could take over the whole ecological niche now being vacated by whales, there will be a lot of it. But despite the popularity of lobsters, prawns and other large crustacea, Daedalus doubts that krill as such will be welcomed by the chef and housewife. It may have to be ground up to a nutritious 'fish-flour', partially hydrolysed to a proteinaceous pabulum, or even fed to chickens. But one way or another, the ever-ingenuous factory-food industry will put it to good use.

David Jones

Obituary

Henry W. Kendall (1926–99)

High-energy physicist and activist

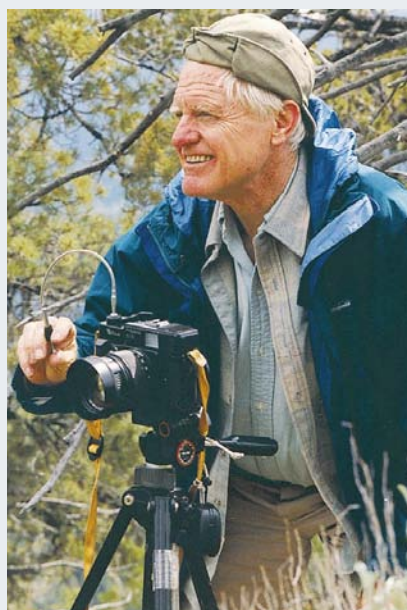
The range of Henry Kendall's accomplishments and avocations is reminiscent of that of a prodigiously energetic and versatile Victorian. He was a key figure in a fundamental discovery in physics, a world-class mountaineer in his younger years, a photographer and diver at the professional level, and an inspiring leader in bringing the concerns of scientists, about the social impact of technology, to the attention of the American body politic.

Kendall was born in Boston in 1926. In his own words, "[I] developed — or had been born with, an active curiosity and interest in things mechanical, chemical and electrical, and do not remember when I was not fascinated with them and devoted to their exploration". Following a mathematics degree from Amherst College, he obtained his doctorate at the Massachusetts Institute of Technology (MIT) in 1954 with a difficult experiment on the spectrum of positronium, the atom formed by an electron and positron, which had recently been discovered by his research advisor Martin Deutsch.

After a postdoctoral fellowship, he joined the Stanford University physics department in 1956. There he met Jerome Friedman and Richard Taylor, who were to become his long-term scientific collaborators. Friedman and Kendall joined the MIT faculty in 1960–61, but maintained a connection with experimental work at Stanford.

The Stanford Linear Accelerator (SLAC), completed in 1966, produced 20 GeV electrons, and provided the means for examining the internal structure of the proton with unprecedented resolution. But it was far from obvious that this technique would yield anything of interest. Protons of the highest available energy were judged by most to be the tool of choice, because at that time it was the widely held view that the proton had no substructure. True enough, Murray Gell-Mann and George Zweig had, independently, already proposed the quark model — the conjecture that the proton is a composite of three spin- $\frac{1}{2}$ objects with charges smaller than that of the proton. But most physicists viewed the model as an intriguing, but abstract, algorithm, and were loath to ascribe physical reality to quarks, let alone a point-like structure.

Kendall was as adventurous in physics as in his other passions, and was



undeterred by this received wisdom. With Friedman and Taylor, he led the MIT–SLAC collaboration in building a pair of gigantic magnetic spectrometers, and he played a key role in the design and construction of the particle detectors and electronics. The team systematically explored highly inelastic electron–proton scattering, which most expected would yield abysmally few data of no interest. What they found, however, were copious data that indicated the proton to be a composite of point-like charges. A deeper analysis suggested by James Bjorken revealed that these objects were loosely bound, and the group later showed they had spin $\frac{1}{2}$. The group's conclusions have, therefore, been compared to Rutherford's realization that the surprisingly large number of back-scattered alpha particles resulting from alpha-particle bombardment of matter implied a point-like nucleus within the atom.

The MIT–SLAC results were a watershed. In combination with subsequent neutrino scattering data from CERN in Geneva, they gave the first clear evidence for the existence of quarks, and provided a foundation for today's remarkably successful theory of the fundamental constitution of matter — the so-called Standard Model of particle physics. In 1990, Friedman, Kendall and Taylor were awarded a Nobel prize for this achievement.

Kendall was as impressed as anyone with the benefits of science. But he came to harbour concerns about the risks that can accompany technological innovation, and developed a growing commitment to

address them. As did many physicists of his generation, he first focused on nuclear weapons: "At the start of the 1960s, troubled by the massive build-up of the superpowers' nuclear arsenals, I joined a group of academic scientists advising the US Defense Department." But he eventually withdrew from classified work and joined colleagues who were founding the Union of Concerned Scientists (UCS). In 1969, UCS mounted a nationwide protest against government policies that "present a major threat to the existence of mankind".

The UCS, which then had no funds or staff, would have been a one-day wonder but for the leadership that Kendall then assumed. His critique of the safety of nuclear power plants brought UCS into high-profile conflict with the Atomic Energy Commission, and contributed to the formation of the Nuclear Regulatory Commission. In the 1980s, UCS was a major player in the Star Wars controversy. By then it had some 100,000 supporters and a sizeable full-time staff.

During the past decade, Kendall's own efforts and those of UCS were guided by his belief that "human beings and the natural world are on a collision course" (the opening words in his 1992 *World Scientists Warning to Humanity*). This declaration, signed by some 1,700 leading scientists from 70 countries, is a synoptic assessment of the factors leading to this collision and a call to action. Kendall ceaselessly sought support for this endeavour, and devoted his own relentless energy, first to the issue of food and water resources, then to climate change and most recently to species extinction.

Kendall had never been on a rope when he first went to Stanford, but within two years he had become a top-flight climber in Yosemite, and a member of expeditions to the Andes. Perhaps his greatest climb was that of the notorious Walker Spur on the Grandes Jorasses in the Alps, with Gary Heming in 1962, the first ascent by Americans and the swiftest by far up to that time. For 50 years he was a serious diver, summer and winter off the New England coast, and as far afield as the Falklands and South Georgia. He died on 15 February 1999, while fresh-water diving in Florida.

He will be remembered as a man who always strove to scale the greatest and most formidable heights, whether in science, in the mountains or in the pursuit of human welfare, with indomitable courage and a cheerful smile for his companions.

Kurt Gottfried

Kurt Gottfried is in the Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853-5001, USA.
e-mail: kg13@cornell.edu

Colour categories in a stone-age tribe

The Dani of Irian Jaya are a stone-age Melanesian people who have provided an empirical basis for the study of cross-cultural perception and cognition¹⁻³. Although they had only two terms for describing colour, the Dani memory for colour seemed to be much like that of modern English speakers. We have investigated another stone-age culture, the Berinmo of Papua New Guinea, for the way in which they categorize colours, but the results do not support the idea that colour categories could be universal.

According to the linguistic relativity hypothesis^{4,5}, which is still influential, we construct our understanding of the world through language. Whorf⁴ famously argued that, to an Eskimo, it would be unthinkable to use the same word for all types of snow because of its wide range of types and different uses.

We investigated colour in a remote, previously unstudied, hunter-gatherer tribe, the Berinmo, which lives on the upper reaches of the Sepik River in Papua New Guinea. When Berinmo subjects were asked to name the 160 colours in the standard Munsell array, they used five basic colour terms⁶. The range and boundaries of these terms showed good intra-subject concordance, and can be seen in Fig. 1 alongside the eight basic chromatic terms in English.

We replicated the Dani experiment with the Berinmo. The accuracy with which they remembered colours bore a striking similarity to the Dani; both groups of Melanesian subjects were very poor at this (9.6 and 7.7 out of 40, respectively). However, statistical analysis showed that, for both studies, the best statistical fit (that with the lowest stress value) was between Melanesian naming and Melanesian memory (Table 1a). This finding is consistent with the linguistic relativity hypothesis, but not with the interpretation of the original study.

The differences between English and Berinmo allow a further critical test of the contrast between colour universals and

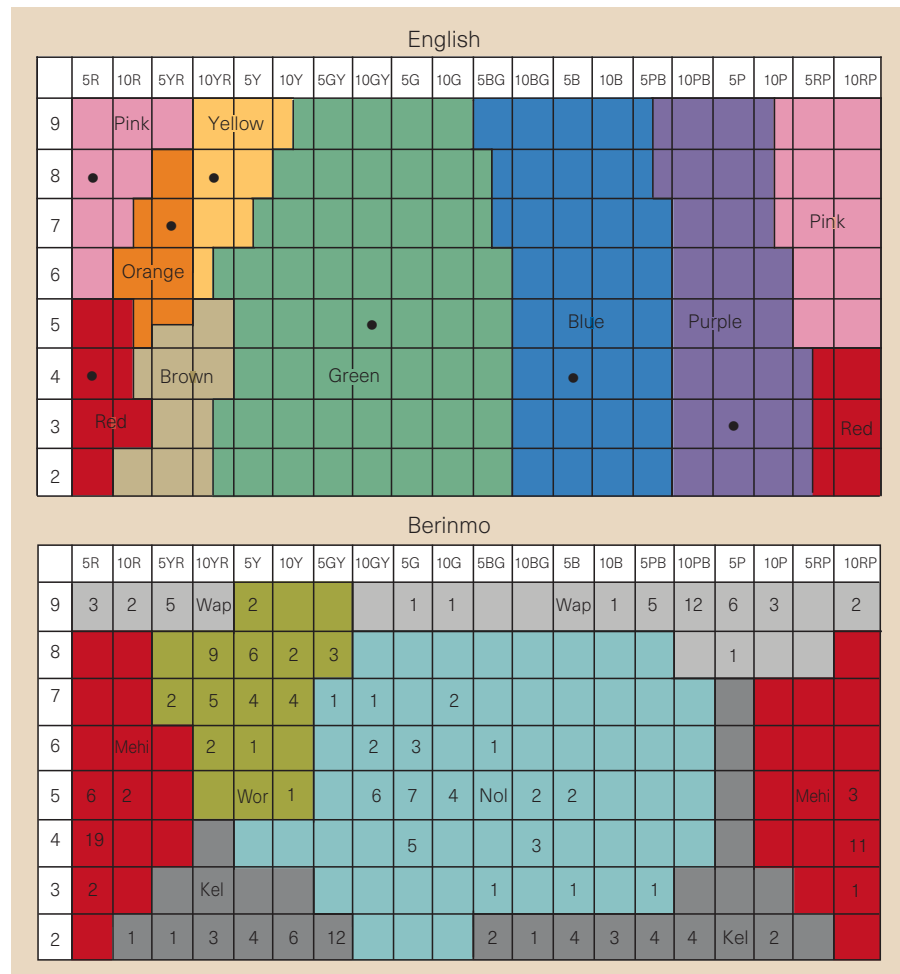


Figure 1 Distribution of English and Berinmo colour names. The Munsell system provides equally spaced samples in three dimensions, but is shown here as a Mercator projection of hue (horizontal axis) against lightness (vertical axis). The colours used to denote colour categories on these Mercator projections are for illustration only. Eight colour terms for English and five for Berinmo are shown. Dots on English naming data represent the position of focal colours². Numbers on the Berinmo naming data represent the number of subjects who designated that colour as best example of the category. R, red; Y, yellow; G, green; B, blue; P, pink.

linguistic relativity. Berinmo does not mark the distinction between blue and green, but it has a colour boundary (between 'nol' and 'wor') in a position that does not exist in English. We investigated categorical effects^{7,8} across both the blue–green and the nol–wor boundaries. We asked subjects to

remember a colour over an interval of 30 seconds^{1,2} and then select the same colour from a pair of similar alternatives. Sometimes the incorrect choice was from the same colour category and sometimes from a different one. We also added a 5-second-interval condition for the Berinmo as they had difficulty remembering blue–green samples for 30 seconds.

English subjects showed the expected advantage for cross-category blue–green decisions but not for nol–wor decisions; Berinmo subjects showed exactly the opposite pattern. The Berinmo showed no sign of a cross-category advantage for blue–green stimuli, but maintained their cross-category advantage for nol–wor stimuli both at 30 seconds and at 5 seconds. These results indicate that categorical perception occurs, but only for speakers of the language that marks the categorical distinction, which

Table 1 Statistical analysis of language data

a Goodness of fit for multidimensional scaling solutions

Dani naming	versus	Dani memory	0.126
Dani memory	versus	US memory	0.161
Berinmo naming	versus	Berinmo memory	0.158
Berinmo memory	versus	English memory	0.256

b Mean trials to criteria in colour-categorization tasks

	Blue vs green	Green1 vs green 2	Nol vs wor	Yellow vs green
English speakers	3.2	5.9	3.8	1.4
Berinmo speakers	11.43	10.57	2.2	3.6

a, Measures of stress (departure from goodness of fit) are shown for comparison of naming and memory data. Low values indicate high goodness of fit. Data for comparisons between US naming and US memory are from ref. 1 and are compared with those from Berinmo and English subjects. In both cases, the fit between naming and memory is better than the fit between memory across language groups. **b**, Mean number of blocks to error-free performance. Categorizations are achieved more rapidly if they are consonant with distinctions made in the language of the subject.

is consistent with the linguistic relativity hypothesis.

If categories always form around natural fault lines in perceptual colour space, it should be relatively easy to learn another language's colour categories. To test this version of the universalist position, we asked English speakers to learn the *no*-*wor* distinction and Berinmo speakers to learn the blue-green and yellow-green distinctions. For comparison, subjects were also asked to categorize stimuli in a manner consonant with the colour names of their own language. In addition, subjects learned a distinction not marked in either language: that between two types of green ('green 1' and 'green 2'). All tasks were made non-trivial by presenting only one stimulus at a time and by the inclusion of marginal examples of each category.

Berinmo subjects found learning to divide colours into green 1 and green 2 no harder than dividing them into blue and green; English speakers found the blue-green task easier. Berinmo subjects found the *no*-*wor* task easier than the yellow-green task, whereas English subjects found the reverse. Tasks in which subjects divided stimuli varying in hue, lightness and saturation into two colour categories are performed better if the division corresponds to a linguistic, rather than a supposed universal, distinction (Table 1b).

Our results from these experiments are consistent with there being a considerable degree of linguistic influence on colour categorization, and place constraints on the type of neuron likely to underpin it. Neurons have been discovered in monkeys that are highly selective to wavelength⁹, to combinations of wavelength and brightness¹⁰ and to colour constancy⁹, but it is unlikely that there are neurons that respond to all examples of a colour category unless their operation is susceptible to linguistic modification.

Jules Davidoff*, Ian Davies†, Debi Roberson*

*Goldsmiths' College, University of London, London SE14 6NW, UK

e-mail: j.davidoff@gold.ac.uk

†Department of Psychology,

University of Surrey,

Guildford, Surrey GU2 5XH, UK

1. Heider, E. R. & Olivier, D. C. *Cogn. Psychol.* **3**, 337–354 (1972).

2. Heider, E. R. *J. Exp. Psychol.* **93**, 10–20 (1972).

3. Hardin, C. L. & Maffi, L. in *Color Categories in Thought and Language* (eds Hardin, C. L. & Maffi, L.) 1–18, 347–372 (Cambridge Univ. Press, 1997).

4. Whorf, B. *Language, Thought and Reality* (MIT Press, Cambridge, Massachusetts, 1956).

5. Saunders, B. A. C. & van Brakel, J. *Behav. Brain Sci.* **20**, 167–228 (1997).

6. Berlin, B. & Kay, P. *Basic Color Terms: Their Universality and Evolution* (Univ. California Press, Berkeley, 1969).

7. Bornstein, M. H. & Korda, N. O. *Psychol. Res.* **46**, 207–222 (1984).

8. Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L. & Rowland, D. *Vis. Cogn.* **3**, 81–117 (1996).

9. Zeki, S. *Neuroscience* **9**, 741–765 (1983).

10. Yoshioka, T., Dow, B. M. & Vautin, R. G. *Behav. Brain Res.* **76**, 51–70 (1996).

Why biodiversity surveys are good value

Article 8 of the Convention on Biological Diversity obliges contracting parties to establish protected areas for conservation. This can be achieved in smaller networks of reserves if their design is based on how well different sites complement one another biologically, rather than on more commonly used criteria, such as species richness or simple availability for acquisition^{1,2}. However, this increase in efficiency³ requires species lists for each candidate site, and obtaining such data can be expensive; for example, a detailed survey of five taxa across 15,000 km² of forest in Uganda took nearly 100 person-years and cost about US\$1 million^{4,5}. Here we ask whether investing in such surveys makes economic sense, or whether conservation agencies would be better advised to continue following more traditional reserve selection procedures, at the cost of having to conserve larger reserve networks.

This trade-off is shown in Fig. 1. Using a simple reserve selection rule (such as buying relatively intact land as it becomes available¹) results in an inefficient reserve network, whose cumulative representation of biodiversity rises only slowly with increasing area. But if a complementarity-based algorithm is used instead, the network needed to achieve a particular conservation goal is reduced by an area *a*. The greater cost of conserving the less efficient network has a present value equal to $[a(x+y/r)]$, where *x* is the mean purchase cost of a unit area of land for conservation, and *y* is the mean cost per unit area of effective maintenance, discounted into the future at an annual rate *r*. In contrast, conducting a high-quality survey costs *zA*, where *A* is the total area of all the candidate sites surveyed, and *z* is the survey cost per unit area. It follows that investing in surveys is good value provided that $zA < [a(x+y/r)]$ or alternatively that

$z < [(a/A)(x+y/r)]$. This therefore sets the upper limit of cost-effective surveys for reserve selection.

There are few published data with which to parametrize this simple model. Conservatively, we suggest that *a/A* (the relative saving in reserve area achieved by tackling complementarity) is commonly at least 5%; it will often be far higher¹. Estimates for *x* and *y*, obtained from a diverse range of sources and expressed in 1990 US\$, are summarized in Table 1. Using appropriate values for *r* (from 5% to 20%), we can then estimate $[(a/A)(x+y/r)]$. Marked variation in land prices⁶ and labour costs means that this upper limit of cost-effective surveys varies enormously but, wherever data are available, this greatly exceeds the likely cost of high-quality surveys. For instance, in Uganda, at *r* = 10%, $[(a/A)(x+y/r)] \approx \$800$ per km², whereas the true value of *z* is less than one-tenth of this, at \$58 per km². Reversing this calculation, detailed inventories would have to yield area savings of less than 0.4% — that is, $a/A < z/[x+y/r]$ — for them not to be worthwhile. This condition is extremely unlikely to be met¹. So, in Uganda, detailed biodiversity inventories represent a very good conservation investment.

In other developing countries, the costs of detailed surveys for reserve selection are similar (*z* ≈ \$65 per km² in Sri Lanka⁷, and *z* ≈ \$135 per km² in Yemen; A. Miller, personal communication), and far less than corresponding values for $[(a/A)(x+y/r)]$, which range from several hundred to a few thousand dollars per km² (Table 1). In developed countries, $[(a/A)(x+y/r)]$ is typically much higher (Table 1) and, although the labour costs of surveys are also high, these are generally offset by the greater availability of existing inventory data. As a result, in the United Kingdom, *z* ≈ \$1,500 per km² (M. Drake and R. Porley, personal communication), which is much less than $[(a/A)(x+y/r)]$. In Australia, on average, $z \ll [(a/A)(x+y/r)]$, at just \$5 per km² (ref. 8). As in some other countries, the Australian mean value masks huge local variation in all costs; nevertheless, in arid areas

Table 1 Estimates of the costs of buying and maintaining nature reserves

	Cost of land purchase, <i>x</i> (\$ per km ²)	Cost of effective maintenance, <i>y</i> (\$ per km ²)	Present value of maintenance, <i>y/r</i> (\$ per km ²)			$(a/A)(x+y/r)$ (assuming <i>a/A</i> = 5%) (\$ per km ²)
			<i>r</i> = 5%	<i>r</i> = 10%	<i>r</i> = 20%	
Uganda	12,628	383	7,660	3,830	1,915	823
Ghana ¹²	45,215	236	4,720	2,360	1,180	2,379
South Africa ¹³	21,896	1,600	32,000	16,000	8,000	1,895
Brazil ^{12,14}	10,776	169	3,380	1,690	845	623
Belize ¹⁵	14,087	350	7,000	3,500	1,750	879
UK	100,523	6,443	128,860	64,430	32,215	11,469
USA ¹²	78,730	2,053	41,060	20,530	10,265	5,990
Australia ^{8,12}	385	359	7,180	3,590	1,795	378

Figures are in 1990 US\$ and often mask considerable local variation. Figures in bold are based on *r* = 10% for developing countries and *r* = 5% for developed countries, but our conclusions are robust for 5% < *r* < 20%. Data were generously provided by T. Butynski, P. Howard, the African Wildlife Foundation, Conservation International, Fauna and Flora International, Kwa Zulu Natal Nature Conservation Service, Royal Society for the Protection of Birds, the Nature Conservancy, and the Western Australia Department of Conservation and Land Management, and cited sources.

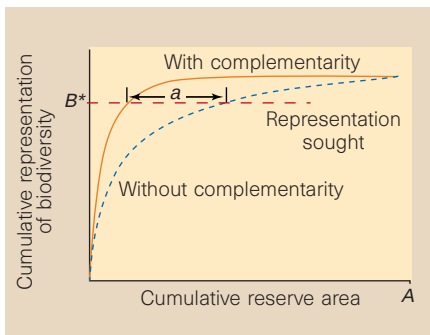


Figure 1 Trade-off in selecting protected areas. Choosing areas without taking between-site complementarity into account results in less efficient reserve networks that need to be larger by a to achieve any given conservation goal (B^*).

where x and y are very low, z is also very low⁸, so that, even locally, detailed surveys are still highly cost-effective.

These conclusions are conservative in ignoring extra costs of inefficient conservation networks (such as opportunity costs and initial expenditure on reserve infrastructure), in using a low figure for a/A , and in neglecting the additional benefits of biodiversity surveys for subsequent reserve management and for a broader understanding of biodiversity.

We are not advocating the widespread adoption of immensely detailed and expensive surveys (such as Costa Rica's All-Taxa Biodiversity Inventory): for the purposes of reserve selection, surveys are worthwhile only if $z < [(a/A)(x + y/r)]$ and if they can be conducted before land acquisition opportunities are lost. Indeed, our model can be modified to establish when short-cut survey techniques that reduce inventory time and costs (but yield species accumulation curves lying between the two lines in Fig. 1) provide savings over fuller inventories.

Similarly, we are not arguing that countries should use a complementarity-based approach simply to identify the smallest area capable of representing each species once: this will often fail to conserve important elements of biodiversity over even the medium term^{9,10}. Instead, selection algorithms should seek several representations of species, and favour clustered reserves over small or scattered ones. Having more ambitious conservation targets such as these is likely to increase the savings achieved by coupling systematic site selection with detailed surveys.

Our results show that, where possible, government agencies and non-governmental organizations should invest in high quality biodiversity inventories before picking protected areas. If they do not, representative reserve systems will be larger than necessary, available resources will be spread more thinly, and conservation objectives will be substantially less likely to be met¹¹.

Andrew Balmford*, **Kevin J. Gaston**

Department of Animal and Plant Sciences,
University of Sheffield, Sheffield S10 2TN, UK
*Present address: Conservation Biology Group,
Department of Zoology, University of Cambridge,
Downing Street, Cambridge CB2 3EJ, UK
e-mail: apb12@cam.ac.uk

1. Pressey, R. L. *Conserv. Biol.* **8**, 662–668 (1994).
2. Williams, P. H. *et al. Conserv. Biol.* **10**, 155–174 (1996).
3. Pressey, R. L. & Nicholls, A. O. *Biol. Conserv.* **50**, 199–218 (1989).
4. Howard, P. C., Davenport, T. & Kigenyi, F. *Oryx* **31**, 253–264 (1997).
5. Howard, P. C. *et al. Nature* **394**, 472–475 (1998).
6. Ando, A., Camm, J., Polasky, S. & Solow, A. *Science* **279**, 2126–2128 (1998).
7. Balmford, A., Jayasuriya, A. H. M. & Green, M. J. B. *Proc. R. Soc. Lond. B* **263**, 1571–1575 (1996).
8. Burbidge, A. A. in *Nature Conservation: Cost Effective Biological Surveys and Data Analysis* (eds Margules, C. R. & Austin, M. P.) 3–6 (CSIRO, Canberra, 1991).
9. Balmford, A., Mace, G. M. & Ginsberg, J. R. in *Conservation in a Changing World* (eds Mace, G. M., Balmford, A. & Ginsberg, J. R.) (Cambridge Univ. Press, in the press).
10. Soulé, M. E. & Sanjayan, M. A. *Science* **279**, 2060–2061 (1998).
11. Leader-Williams, N. & Albon, S. D. *Nature* **336**, 533–535 (1988).
12. James, A. N., Green, M. J. & Paine, J. R. *Financial Indicators for Biological Diversity Conservation: A Global Analysis of Protected Area Investment* (World Conservation Monitoring Centre, Cambridge, 1997).
13. Anon *Oryx* **24**, 67 (1990).
14. Anon *Parks* **2**, 52 (1991).
15. *Nature Conservancy Parks in Peril Source Book* (The Nature Conservancy, Arlington, VA, 1995).

How the 'terror crocodile' grew so big

Deinosuchus is a giant crocodylian from the Late Cretaceous period of North America. It was 8 to 10 metres long and weighed between 2,500 and 5,000 kg, three to five times more than the largest crocodiles alive today. How *Deinosuchus* attained sizes to rival its dinosaurian contemporaries, on which it undoubtedly preyed, has remained a mystery. Did it exhibit accelerated growth

rates, like its dinosaurian cousins¹, or did it simply maintain primitive reptilian rates for decades (as was once proposed to explain gigantism in dinosaurs²)? We find that growth indices from *Deinosuchus* skeletons reveal rates comparable to those of smaller crocodylian taxa, indicating that the gigantic proportions were attained by prolonging development.

We reconstructed growth patterns in *Deinosuchus* by coupling age and size estimates throughout development, and determined longevity by counting growth rings in dorsal osteoderms from several individuals from the Campanian Judith River Formation of Montana and the Aguja Formation of Texas³. Such rings form annually throughout the skeleton in extant members of Crocodylia and in proximate outgroup clades (such as Lepidosauria and Chelonio^{3,4}). We based estimates of total length for adult *Deinosuchus* on a large mandibular ramus from Texas and regression curves for the American alligator (*Alligator mississippiensis*) and saltwater crocodile (*Crocodylus porosus*)^{5,6}. We used proportions of linear increase per growth ring as standardized measures to estimate annual increases in total body length throughout ontogeny, and conducted comparable analyses on the osteoderms of large individuals from clades closely related to *Deinosuchus*⁷ (Fig. 1). In Fig. 2a, length is plotted against age for the extinct taxa; growth curves for large extant crocodylians are also shown. We have made osteohistological characterizations of the jaws, ribs, vertebrae and long bones of *Deinosuchus*, and assessed relative somatic growth rates¹.

We estimate longevity in the *Deinosuchus* specimens to be 50 and 51 (± 2) years (Fig. 2b). Extant crocodylians rarely live this long in the wild⁸, and none of the fossil specimens from related lineages attained such ages (Fig. 2a). We estimated

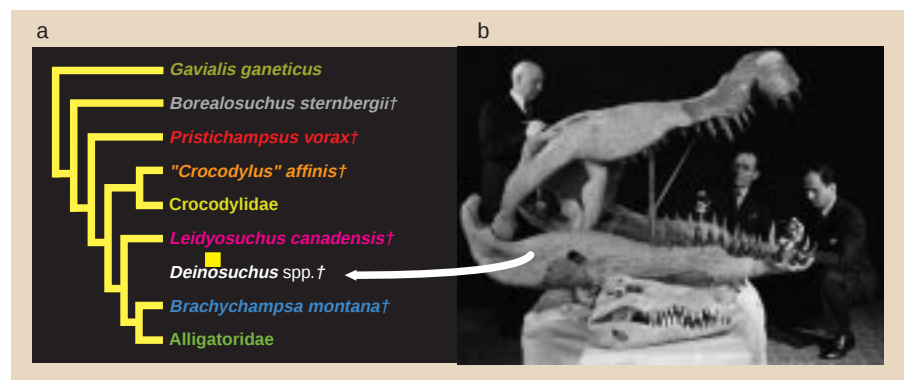


Figure 1 *Deinosuchus* and its phylogenetic position. **a**, Phylogenetic hypothesis for the Crocodylia showing the relationship of *Deinosuchus* to other clades. The cladogram is based on maximum-parsimony analysis of 164 morphological characters from both fossil (†) and living taxa. **b**, The size of an adult *Deinosuchus* skull relative to a specimen of a large extant crocodile. Pictured from left to right are B. Brown, R. T. Bird and E. M. Schlaikjer (negative no. 318651, courtesy of the Department of Library Services, American Museum of Natural History). Although the overall size of the *Deinosuchus* reconstruction is generally correct, recent finds show the skull to have had a broader snout, giving it a more alligator-like appearance. Growth rings in pre-caudal dorsal osteoderms were studied from extinct non-gavialoid crocodylians in the phylogenetic bracket.

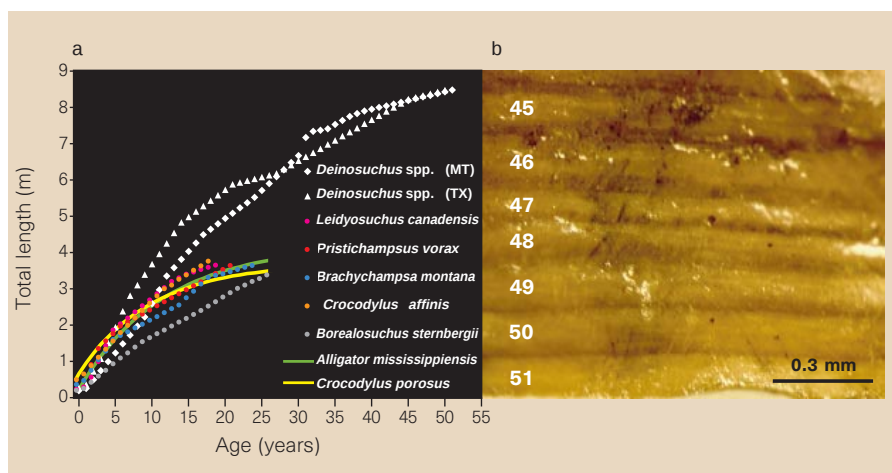


Figure 2 Growth rates of crocodylians and osteoderm growth rings. **a**, Total length against age in non-gavialoid crocodylians. The data for individuals from extinct clades (symbols; specimen numbers are available from the authors on request) are based on growth line counts in dorsal osteoderms extrapolated to total body length. The idealized growth curves for large extant crocodylians (lines) are from capture-recapture analyses from hundreds of individuals^{10,11}. The similar patterns between the two data sets for specimens of comparable sizes indicate that the methods used provide reasonable estimates of crocodylian growth. The *Deinosuchus* growth curves show a continuation of ancestral juvenile growth rates past 10 years of age and presumably greater longevity. **b**, Lamellar bone growth rings from the ventral surface of an osteoderm representing the final seven years of growth (years 45 to 51) in a specimen of *Deinosuchus* from the Aguja Formation of Texas.

the total length for the *Deinosuchus* specimens to be 8.43 to 9.10 metres. Plots of length against age for *Deinosuchus* outgroups show that the primitive character state for growth in non-gavialoid crocodylians was characterized by rapid linear increases early in development, with rates typically declining by the first decade after hatching (Fig. 2). *Deinosuchus* showed similar rates (about 0.3 metres per year) to other crocodylians from the phylogenetic bracket during the first five to ten years of life, but maintained these juvenile growth rates for several decades (Fig. 2).

The primary histological structure of *Deinosuchus* skeletal elements supports the interpretation that primitive developmental growth rates were retained. Bones from the giant crocodylian showed slowly deposited lamellar-zonal bone tissue typical of non-gigantian crocodylians (the ancestral pattern in reptiles). Because patterns of bone fibre packing correlate with osseous deposition rates¹, it seems that bone formation was not accelerated appreciably. This is in contrast to the skeletons of dinosaurs, which show evidence of derived, accelerated rates in the form of rapidly deposited fibro-lamellar bone¹.

The evolution of increased metabolic rates in dinosaurs is believed to have facilitated the evolution of gigantism by enabling them to build their skeletons swiftly using fibro-lamellar bone¹. *Deinosuchus* achieved the same outcome, but it took much longer. Dinosaurs of similar size to *Deinosuchus*, such as hadrosaurs ('duck-billed dinosaurs'), reached adult size in only seven to eight years⁹, whereas the giant crocodylian

required more than 35 years. We believe that the retention of an ectothermal physiology constrained *Deinosuchus* to the deposition of slow-forming somatic tissues (such as lamellar bone) throughout development, necessitating a greater developmental time to reach dinosaurian proportions.

Gregory M. Erickson*†, Christopher A. Brochu‡

*Department of Integrative Biology and Museums of Vertebrate Zoology and Paleontology, University of California, Berkeley, Berkeley, California 94720-3140, USA

‡Department of Geology, Field Museum of Natural History, Lake Shore Drive at Roosevelt Road, Chicago, Illinois 60605, USA

†Present address: Biomechanical Engineering Division, Mechanical Engineering Department, Stanford University, Stanford, California 94305, USA
e-mail: erickson@bones.stanford.edu

1. Ricqlès, A. de in *A Cold Look at the Warm-blooded Dinosaurs* (eds Thomas, R. D. K. & Olson, E. C.) 103–139 (Westview, Boulder, Colorado, 1980).
2. Case, T. J. *Paleobiology* **4**, 320–328 (1978).
3. Hutton, J. M. *Copeia* **1986**, 332–341 (1986).
4. Castanet, J., Francillon-Vieillot, H., Meunier, F. J. & Ricqlès, A. de in *Bone Vol. 7* (ed. Hall, B. K.) 245–283 (CRC, Boca Raton, Florida, 1993).
5. Woodward, A. R., White, J. H. & Linda, S. B. *J. Herpetol.* **29**, 507–513 (1995).
6. Greer, A. E. *J. Herpetol.* **8**, 381–384 (1974).
7. Brochu, C. A. *Phylogenetic Systematics and Taxonomy of Crocodylia*. Thesis, Univ. Texas, Austin, 1997.
8. Grenard, S. *Handbook of Alligators and Crocodiles* (Krieger, Malabar, Florida, 1991).
9. Ricqlès, A. de, Horner, J. R. & Padian, K. J. *Vert. Paleontol. Abstr.* **18**, 72 (1998).
10. Chabreck, R. H. & Joanen, T. *Herpetologica* **35**, 51–57 (1979).
11. Webb, G. J. W. et al. *Aust. Wildl. Res.* **5**, 385–399 (1978).

Localization or classical diffusion of light?

Wiersma *et al.*¹ have reported near-infrared optical transmission and coherent backscattering data from strongly scattering slabs of micrometre-sized semiconductor particles. Their optical transmission was much weaker, and the angular shape of their coherent backscattering more rounded, than would be expected for classical diffusive light propagation without absorption. The authors interpret this as evidence for the onset of strong localization of light, but we find that their data can be explained by classical diffusion combined with reasonable amounts of absorption. Moreover, the turbidities of their samples are much lower than those given in ref. 1 and are comparable to samples with classical transport properties. We therefore question whether their samples are in fact close to the proposed localization transition.

In diffusive transport through disordered non-absorbing media in which $\lambda \ll l^* \ll L$, we have $T \propto l^*/L$, where $\lambda = 2\pi/k$ is the wavelength, l^* is the transport mean free path, T is the optical transmission, and L is the slab thickness. For $kl^* \rightarrow 1$, the scaling theory of localization predicts that $T \propto L^{-2}$, whereas in the regime of strong localization ($kl^* \leq 1$), $T \propto \exp[-L/l_{\text{loc}}]$, where l_{loc} is the localization length. In classical diffusion ($kl^* \gg 1$) with absorption, T crosses over from $T \propto L^{-1}$ at small L to $T \propto \exp[-L/L_a]$, where L_a is the absorption length, at large L . Thus, it is difficult to distinguish localized light from absorbed light solely on the basis of measured average light intensities.

Wiersma *et al.*¹ used powders of gallium arsenide produced by grinding. No significant distortion of the bandgap was seen, but this does not exclude absorption for wavelengths inside the gap, because contributions from surface states, dangling bonds, defects, internal stress or impurities, for example, are likely to increase upon grinding. For the 1- μm particles, the coherent-backscattering cones were found to be substantially flatter close to backscattering than for non-absorbing samples, but were not much wider at larger angles (Fig. 1a). Figure 1a compares data from ref. 1 with data from a dense powder of non-absorbing TiO_2 particles². Because the cone width scales with kl^* and the TiO_2 sample has $kl^* \approx 5.8$ (ref. 2), the comparable widths suggest that $kl^* \approx 5.0 \pm 1$ ($l^* \approx 0.8 \mu\text{m}$) for this GaAs sample, in contrast to $kl^* \approx 1.5$, as given in ref. 1. The latter value was obtained by fitting the low-angle (≤ 10 mrad) slope in the region of strong rounding. This provides unphysically

small l^* values (also seen from the dashed line in Fig. 4b of ref. 1), making $T(L)$ much lower than the data, whereas the classical T values must always lie above any data from a localizing — or absorbing — sample. For more quantitative comparison, we slightly rescaled the angular axis of the GaAs cone (Fig. 1a). The agreement becomes striking if we introduce an absorption length $L_a = \sqrt{(l^*l_a/3)} \approx 8 \mu\text{m}$, or $l_a \approx 300 \mu\text{m}$ for TiO_2 using classical diffusion theory. Note that identical results are obtained from numerical simulations. We conclude that kl^* is 5 ± 1 rather than 1.5, and that the shape of the GaAs cone can be explained by classical diffusion including absorption.

Next we analysed the transmission of the same sample¹. As $l^* \geq 0.5 \mu\text{m}$, the classical transmission without absorption lies above the experimental values for all cases. Figure 1b shows that the T data can be described by diffusion theory with absorption over the whole L range; l^* and l_a found by fitting are very close to values obtained from coherent-backscattering cones, emphasizing the consistency of our analysis. Finally, for the smallest particle size ($\sim 0.3 \mu\text{m}$), an exponential decay with

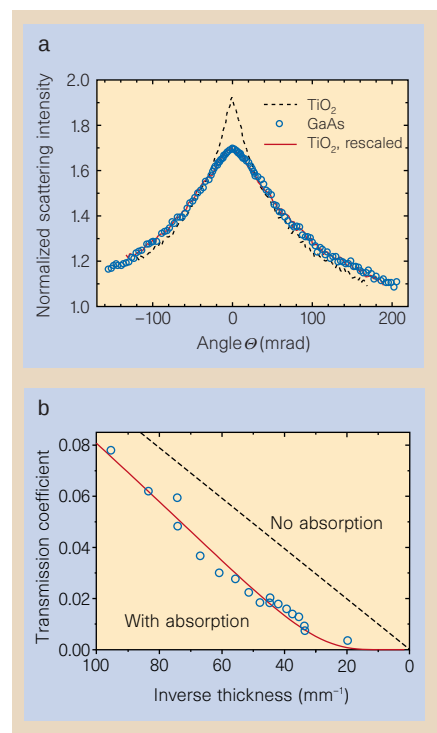


Figure 1 Optical properties of very turbid powders. a, Comparison of coherent backscattering cones for GaAs powder (Fig. 3b of ref. 1) and for pure TiO_2 (ref. 2). The red line represents data for TiO_2 that include absorption transformed according to $\theta \rightarrow \theta - \theta_a^2$, where $\theta_a = 1/(kl_a) = 21 \text{ mrad}$, and rescaled by $\theta \rightarrow 1.13 \times \theta$. b, Plot of transmission T of GaAs powder (circles) of average particle size $1 \mu\text{m}$ against inverse sample thickness L^{-1} (Figs 4b and 5b of ref. 1). The red line is derived from diffusion theory with absorption $T(L) = (L_a / \gamma l^*) \sinh^2 (\gamma l^* / L_a) / \sinh(L / L_a)$, $\gamma = 5/3$, $l^* = 0.59 \mu\text{m}$, $L_a = 8.9 \mu\text{m}$. For the dotted line, $L_a \rightarrow \infty$.

decay length of $\sim 4.3 \mu\text{m}$ was reported¹. This is a reasonable L_a value, given the expected stronger absorption and smaller l^* of this sample.

In conclusion, we believe that the results of Wiersma *et al.* cannot be considered as unambiguous evidence for the existence of the critical regime, or Anderson localization, of light. Further investigation is needed to quantify the role of absorption.

Frank Scheffold, Ralf Lenke, Ralf Tweert, Georg Maret

Fakultät für Physik, Universität Konstanz,
78457 Konstanz, Germany
e-mail: georg.maret@uni.konstanz.de

1. Wiersma, D. S., Bartolini, P., Lagendijk, A. & Righini, R. *Nature* **390**, 671–673 (1997).
2. Wiersma, D. S., van Albada, M. P., van Tiggelen, B. A. & Lagendijk, A. *Phys. Rev. Lett.* **74**, 4193–4196 (1995).

Wiersma *et al.* reply — Scheffold *et al.* have compared different data sets from our group and suggest that our evidence for localization is not conclusive and that the role of absorption should be characterized further. We believe that their analysis is misleading, and that our conclusions about localization are still valid.

Scheffold *et al.* compared our data on coherent backscattering from GaAs powders¹ with our previously published data² on coherent backscattering from TiO_2 . From a comparison of the widths of the cones, they estimate the inverse of the scattering strength, kl^* , of our GaAs samples, where k is the wavevector of the light and l^* is the mean free path. However, the GaAs data were recorded using linearly polarized light, whereas the TiO_2 data were recorded with circularly polarized light². This difference is fundamental, as linearly polarized light inherently gives a smaller enhancement factor (the top of the cone is lower) than circularly polarized light. To overlay our two data sets, Scheffold *et al.* scaled the y-axis of one of the data sets and shifted the zero of its y-axis. The results are misleading, however, because shifting the zero also changes the full width at half maximum of the cone, and so gives the wrong value of kl^* , because the width is inversely proportional to kl^* .

To illustrate the inconsistency of this approach, we repeated the procedure used by Scheffold *et al.*, but keeping the TiO_2 instead of the GaAs data fixed, and scaling and shifting the y-axis of the GaAs data. The result indicated that the backscattering cone from GaAs is three times wider than that from TiO_2 , which corresponds to a scattering strength of the GaAs samples of $kl^* = 1.7$, instead of 5.0. This shows that manipulating the y-axis leads to strongly varying and thereby inconsistent conclusions about the same experiment.

A strong argument against absorption, for example from surface states, is that the

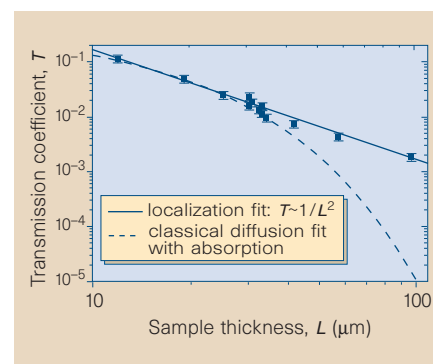


Figure 2 Comparison of scaling theory and classical diffusion using transmission coefficients of a GaAs sample. Scaling theory, solid line; classical diffusion with absorption, broken line. Average particle diameter is $1 \mu\text{m}$, mean free path l^* is $0.59 \mu\text{m}$, and absorption length L_a is $9.6 \mu\text{m}$. Data show the L^{-2} behaviour characteristic of the localization transition, and cannot be fitted as classical diffusion with absorption.

shape of the bandgap does not change upon grinding (see the temperature measurements in Fig. 1 in ref. 1). Additional evidence that excludes absorption is the observation that the top of the backscattering cone remains triangular while the surrounding region becomes round. This is exactly the behaviour expected at the localization transition, whereas in the case of absorption, the top of the cone would never be triangular. However, Scheffold *et al.* have replotted our data in Fig. 1a in such a way that this triangular shape is obscured by the use of larger symbols.

Scheffold *et al.* use classical diffusion theory with absorption to fit our transmission data (Fig. 1b). An absorption fit yields a χ^2 of 0.033, whereas our fit with scaling theory (that is, with the desired L^{-2} behaviour instead of exponential decay) yields a better value of 0.0053. Our independent, new transmission measurements using GaAs powders (Fig. 2) show L^{-2} behaviour over two orders of magnitude, confirming again the scaling behaviour at the localization transition and the absence of absorption.

Diederik S. Wiersma*, Jaime Gómez Rivas†, Paolo Bartolini*, Ad Lagendijk†, Roberto Righini*

*European Laboratory for Non-Linear Spectroscopy, Largo E. Fermi 2,
50125 Florence, Italy

†Van der Waals-Zeeman Institute,
University of Amsterdam,
1018 XE Amsterdam, The Netherlands
e-mail: wiersma@lens.unifi.it

1. Wiersma, D. S., Bartolini, P., Lagendijk, A. & Righini, R. *Nature* **390**, 671–673 (1997).
2. Wiersma, D. S., van Albada, M. P., van Tiggelen, B. A. & Lagendijk, A. *Phys. Rev. Lett.* **74**, 4193–4196 (1995).

editorial note: This exchange has been subject to unusual delays during the editorial process: the comment by Scheffold *et al.* was submitted on 22 April 1998 and the reply from Wiersma *et al.* on 7 December 1998.

Simulating the motion of gas bubbles in a liquid

Understanding the motion of gas bubbles in a liquid is a problem of both scientific and engineering importance. About 500 years ago, Leonardo da Vinci¹ summarized his observations on the motion of air bubbles in a liquid: "The air that submerged itself with the water... returns to the air, penetrating the water in sinuous movement, changing its substance into a great number of forms... it never spreads itself out from its path except to the extent to which it avoids the water which covers it". We have attempted to simulate the motion of single gas bubbles in a liquid using the volume-of-fluid (VOF) technique, which allows us to describe the complex bubble dynamics using only the fluid phase properties as inputs.

The VOF model resolves the transient motion of the gas and liquid phases using the Navier–Stokes equations, and accounts for the topology changes of the gas–liquid interface induced by the relative motion between the dispersed gas bubble and the surrounding liquid. The finite-difference VOF model uses a donor–acceptor algorithm² to obtain and maintain an accurate and sharp representation of the gas–liquid interface. The VOF method defines a fractional volume or 'colour' function $c(x, t)$, which is a function of position vector x and time, t . The colour function indicates the fraction of the computational cell that is filled with liquid, and varies between 0, if the cell is completely occupied by gas, and 1, if it consists only of liquid. The location of the bubble interface is tracked in time by solving a balance equation for this function, $(\partial c(x, t)/\partial t) + \nabla \cdot (uc(x, t)) = 0$, where u is the velocity vector.

The liquid and gas velocities are assumed to equilibrate over a very small distance, and essentially $u_k = u$ at the bubble interface, where the subscript k corresponds to either liquid or gas. The mass and momentum conservation equations are considered to be homogeneous. The continuum surface force model³ is used to model the force arising from surface tension acting on the gas–liquid interface. In this model, the surface tension is modelled as a body force F_{st} , which is non-zero only at the bubble interface and is given by the gradient of the colour function, $F_{st} = \sigma \kappa(x) \nabla c(x, t)$, where σ is the surface tension and $\kappa(x)$ is the local mean curvature of the bubble interface $\kappa(x, t) = -\nabla \cdot (n/|n|)$, where n is the vector normal to the bubble interface.

All our simulations were carried out in a rectangular column using a uniform two-dimensional cartesian-coordinate grid. The

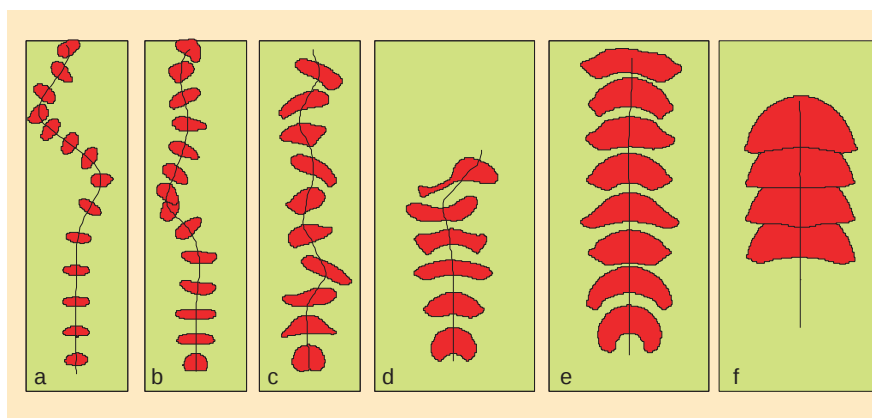


Figure 1 Two-dimensional volume-of-fluid simulations of the rise trajectories of bubbles in the size range 4–20 mm. The bubble sizes are: a, 4 mm; b, 5 mm; c, 7 mm; d, 9 mm; e, 12 mm; f, 20 mm. In all simulations, the gas phase was air (density, 1.29 kg m^{-3} ; viscosity, $1.7 \times 10^{-5} \text{ Pa s}$) and the liquid was water (density, 998 kg m^{-3} ; viscosity, 10^{-3} Pa s ; surface tension, 0.072 N m^{-1}). The Morton number⁶ for all simulations was 2.63×10^{-11} . The Eötvös numbers⁶ for the various bubble sizes were: a, 2.2; b, 3.4; c, 6.7; d, 11.0; e, 19.6; f, 54.5. Animations of the simulations can be seen on our website (http://ct-cr4.chem.uva.nl/single_bubble/).

front of the two-dimensional rectangular grid is formed by the x – z plane. The front and rear faces of the column are modelled as symmetry planes. At the two walls, the no-slip boundary condition is imposed. The column is modelled as an open system, so the pressure in the gas space above the initial liquid column is equal to the ambient pressure (101 kPa). Hybrid differencing was used for the convective terms in the equations, and upwind differencing was used for time integration. The time steps used in the simulations were usually 0.0003 s or smaller. To counteract excessive smearing of the liquid–gas interface by numerical diffusion, a surface-sharpening routine was invoked. This routine identifies gas and liquid on the 'wrong' side of the interface, and moves it back to the correct side, while conserving the volume of the respective phases.

To avoid dissolution of the bubble by surface sharpening, we ensured that the area of each bubble encompassed a few hundred cells. For simulations of bubbles smaller than 12 mm, we found that the rich dynamic features could be captured only by allowing at least 800 grid cells per bubble cross-section, resulting in grid sizes as small as 0.125 mm for the smallest bubble size we could simulate with adequate accuracy. To simulate the rise of spherical cap bubbles, typically found with sizes above 17 mm, a grid size of 1 mm was found to be small enough; in this case, there were more than 300 grid cells per bubble cross-section.

All simulations were carried out using the parallel version of CFX 4.1c (AEA, Harwell, UK) running on a Silicon Graphics Power Challenge machine with six R8000 processors. This package is a finite volume solver that uses body-fitted grids, which were non-staggered, and all variables were evaluated at the cell centres. We used an improved version of the Rhie–Chow algorithm⁴ to calculate the velocity at the cell

faces, and the SIMPLEC algorithm⁵ to obtain the pressure–velocity coupling. Simulating the rise of a bubble 7 mm in diameter for 0.75 s in a column 0.025 m wide and 0.09 m high, involving 144,000 grid cells, required about two weeks of computer processing time.

Typical bubble trajectories are shown in Fig. 1. Bubbles of 4 and 5 mm in diameter meander up the column with large lateral movements, conforming with the observations of Leonardo da Vinci. The 7-mm bubble oscillates from side to side when moving up the column. Bubbles 8 and 9 mm in diameter behave rather like jellyfish. The 12-mm bubble moved like a bird flapping its wings; we could find no mention of this type of bubble motion in the literature, but we were able to observe it in our experiments.

The two-dimensional VOF simulations give velocity values for the rising bubbles that are roughly half those measured experimentally in cylindrical columns. A three-dimensional simulation would be required to determine rise velocity quantitatively and to reproduce precisely the experimental observations of Da Vinci. Such a simulation would require a grid with a few million grid cells and the use of a massively parallel computer.

R. Krishna, J. M. van Baten

Department of Chemical Engineering,
University of Amsterdam,
Nieuwe Achtergracht 166,
1018 WV Amsterdam, The Netherlands
e-mail:krishna@chemeng.chem.uva.nl

- Leonardo da Vinci CD-ROM (Corbis, 1996).
- Hirt, C. W. & Nichols, B. D. *J. Comput. Phys.* **39**, 201–225 (1981).
- Brackbill, J. U., Kothe, D. B. & Zemarch, C. *J. Comput. Phys.* **100**, 335–354 (1992).
- Rhie, C. M. & Chow, W. L. *AIAA J.* **21**, 1525–1532 (1983).
- Van Doormal, J. & Raithby, G. D. *Num. Heat Transfer* **7**, 147–163 (1984).
- Clift, R., Grace, J. R. & Weber, M. E. *Bubbles, Drops and Particles* (Academic, San Diego, 1978).

Tale of a stigma redeemed

The White Death: A History of Tuberculosis

by Thomas Dormandy

Hambledon: 1999. 433 pp. £25

Richard Davenport-Hines

The Texan local historians Sarah and Thad Sitton in their *Austin's Hyde Park* (Pecan, 1991) quote a local woman reminiscing about her sister's death from tuberculosis in the 1920s. The sister felt she had the symptoms, but after being tested by her physician, was assured by him that the results were negative. Still unconvinced, she arranged her own test, which was reported positive. When she confronted her physician with this discrepancy, he said, "I know, child, but yours is such a nice family that I hated for you to have the disgrace." Such incidents show the shame with which tuberculosis was long surrounded.

Mortality rates were as dire as its social stigma. Tuberculosis has taken an estimated one billion lives in the nineteenth and twentieth centuries. Between 1800 and 1870 the disease accounted for one out of every five deaths in the United States. It was the chief cause of death there until 1909. Then, as now, it flourished among the poor, the overcrowded, the malnourished and the disadvantaged. It struck at young adults. Tuberculous meningitis was one of the most acutely feared forms of the disease, often killing children within 48 hours, or leaving them permanently brain-damaged.

During the high point of the disease in the Western world there were two main sources of infection: bovine tuberculosis was caught by drinking milk from infected cattle; human tuberculosis was spread by its victim coughing up heavily infected phlegm to be inhaled as droplets or as dust when dry. It has been estimated that the prevalence of the disease exceeds the mortality 20-fold; resistance and immunity to the disease may be natural or acquired.

Physicians interested in the aetiologies of tuberculosis accordingly searched in the characters as much as the circumstances of their patients. The fever, toxæmia and restlessness that were characteristic of the late stages of tuberculosis enhanced the creativity of John Keats, Anton Chekhov, D. H. Lawrence and other authors whose terminal illnesses are sensitively analysed by Thomas Dormandy in *The White Death*.

Many manifestations of the disease were less acceptable. It was believed to excite sexual adventurousness in women, for example, and its enervating effects made its victims unsatisfactory employees. As a result of such prejudiced social judgements, medical treatments often reflected the belief that tubercu-



Out in the open: the shame originally associated with tuberculosis had been lost by the mid-1900s.

losis was non-contagious, hereditary and an indication of bad character. Although tuberculosis was finally recognized as infectious in 1865, the condemnatory attitudes to many patients persisted long into the twentieth century. This perpetuated the harshness of treatment by medical authorities, the social ostracism of patients, and hence the shame echoed by the Austin doctor.

At the end of the nineteenth century, consumption was renamed tuberculosis; laboratory testing, microscopes and new chemotherapies rightly endowed medical science with new prestige. Strict rules to protect the public from contagious tubercular patients were introduced in Britain and the United States from the 1880s. Some measures — notably compulsory notification of tubercular patients to official bodies by physicians — seldom improved management of the disease and accentuated patients' problems.

As Dormandy describes, there was also a strong movement to build sanatoria for tubercular patients in clean-aired country districts. These sanatoria were intended both to improve patients' chances of survival and to isolate them from the healthy population. Sanatorium discipline was unrelentingly coercive in state-funded institutions. While sanatorium keepers treating rich invalids advocated rest therapy, the superintendents of institutions with working-class inmates imposed strict regimentation and outdoor labour: patients were set to work gardening and tree-felling, which visibly improved the hospital grounds, but less obviously improved the patients.

In the developed world after 1945 mass inoculations with the BCG vaccine, the introduction of mass screenings by radiography and the development of antibiotics seemed to promise the eradication of the disease; but in the poorer nations tuberculosis has remained a leading cause of death. The concluding sections of Dormandy's book provide grim reminders that tuberculosis has now returned in drug-resistant strains and has resurged among the socially disadvantaged in the United States and elsewhere.

Tuberculosis has been the subject of many fine historical books. In the English-speaking world alone, F. B. Smith's *The Retreat of Tuberculosis 1850–1950* (Croom Helm, 1988) provides an excellent overall account of the incidence, aetiologies, changing fashions in treatment and governmental responses to the disease. The personal, social and cultural consequences are treated in Linda Bryder's *Below the Magic Mountain* (Oxford University Press, 1988) for twentieth-century British people, and for North Americans in Barbara Bates's *Bargaining for Life* (University of Pennsylvania Press, 1992) and Sheila Rothman's *Living in the Shadow of Death* (Johns Hopkins University Press, 1992). By studying such artefacts as spittoons, Katharine Ott's *Fevered Lives* (Harvard University Press, 1997) traces how invalidism became a commodity and how medical consumerism developed from 1870. Frank Ryan's *The Forgotten Plague* (Little, Brown, 1993) recounts the achievements of great scientists such as the German bacteriologist Robert Koch, who isolated the bacillus of tuberculosis in 1882, the German pathologist Gerhard Domagk

and the Russian-born Selman Waksman of Rutgers University, whose team in 1943 isolated streptomycin, which proved a potent antibiotic treatment against tuberculosis. Each of these books combines meticulous scholarship with compassion and imagination. This distinguished collection of medical history has now been joined by Dormandy's *The White Death*.

Dormandy is a consultant pathologist who has for many years collected references to the history of tuberculosis, and the impact of the disease on society, the arts and literature. His approach is erudite, witty, humane and authoritative. He has some masterful character portraits of physicians and scientific researchers, and has studied his text with arresting maxims: he writes of Koch that "his postulates established the science of medicine as the equal of the art of medicine". There are many pleasing idiosyncrasies and digressive footnotes; but best of all Dormandy provides a truly international survey which repeatedly cites examples that will be unfamiliar to English-language speakers. Although there are some amateurish repetitions and evidence of obtrusively inadequate proofreading, *The*

White Death can be enthusiastically recommended to all those interested either professionally in tuberculosis or recreationally in the cultural history of medicine. □

Richard Davenport-Hines is at 51 Elsham Road, London W14 8HD, UK.

... nor any drop to drink

Fresh Water

by E. C. Pielou

University of Chicago Press: 1998. 274 pp.

\$24.00, £19.25

The World's Water 1998–1999: The Biennial Report on Freshwater Resources

by Peter H. Gleick

Island: 1998. 306 pp. \$29.95, £24.95 (pbk)

Daniel Hillel

In many ways, E. C. Pielou's comprehensive and comprehensible book describing the cycle of water in the terrestrial environment (especially as it is influenced by human inter-

vention) is reminiscent of Luna Leopold's classic *Water: A Primer*, published some 25 years ago. *Fresh Water* is evidently intended to teach lay readers about the crucial issues posed by the dwindling supplies and deteriorating quality of fresh water available for human use. Her writing is didactic and definitive, in places even charming, and is buttressed by clear illustrations, avoiding rigorous mathematical formulations. Especially enlightening are the explanations of such complex phenomena as hydraulic jumps, the transition from laminar to turbulent flow, water movement in aquifers, the nature of karst and the tendency of rivers to meander.

Considering the enormous breadth and complexity of the topic, however, it is not surprising to discover a few flaws. Simplistic statements occur here and there. One is the characterization of clay as 'rock flour', when in fact clay generally consists of altered ('secondary') minerals (for example, aluminosilicates) with very different physico-chemical attributes from those of the original ('primary') minerals of the parent rocks.

In retrospect chosen by Anton Zeilinger

Albert Einstein:

Philosopher—Scientist

edited by Paul Arthur Schilpp

"Here I sit in order to write, at the age of 67, something like my own obituary." With these words Albert Einstein starts his only existing autobiography, his "Autobiographical Notes". Around the middle of the twentieth century the philosopher Paul A. Schilpp published a series of books under the heading "The Library of Living Philosophers"; each contains essays written by one leading philosopher — such as Karl Popper, Bertrand Russell or Jean-Paul Sartre — and by some of his peers.

Einstein's "Autobiographical Notes", written in German and translated into English by Schilpp, with both versions in the book, are among the most outstanding intellectual reflections ever written. He tells us that "Even when I was a fairly precocious young man the nothingness of the hopes and strivings which chase most men restlessly through life came to my consciousness with considerable vitality". But he then goes on to ask the humble question, "What, precisely, is 'thinking'?" He agrees with the philosopher David Hume that certain concepts, like causality, cannot be deduced, but he disagrees that the specific concepts chosen by Immanuel Kant are indispensable, preferring to call them "freely chosen conventions".

Einstein discusses many issues. He complains about university education. "It is, in fact, nothing short of a miracle that the modern methods of instruction have not yet entirely strangled the holy curiosity of inquiry." One can only regret

that most countries' education systems have taken the road towards less rather than more freedom as demanded by Einstein.

Later on, Einstein gives us many details of how he was led to his theories of relativity and their fundamental ideas, mentioning most prominently the influence of Ernst Mach. It is very moving when he tells us that by the age of 16 he had realized that if one could proceed alongside a beam of light with the same speed, one would see impossible configurations of electric fields.

Parts of the book have already become classics, for example where Einstein discusses his views of quantum mechanics and where Niels Bohr documents his "Discussions with Einstein about epistemological problems in atomic physics". The latter is one of the most momentous debates that ever took place. Einstein, while he admits quantum theory is "the most successful physical theory of our period", nevertheless concludes that it "offers no useful point of departure for future development". He reaches that conclusion by analysing what he believes physics should be about. For him, "Physics is an attempt conceptually to grasp reality as it is thought independently of its being observed. In this sense one speaks of 'physical reality'."

Einstein carefully analyses the situation of two quantum-mechanically correlated, so-called entangled, systems. He points out that observation of one system immediately changes the quantum state of the other regardless of how far apart they are. Yet, as Einstein says, "the real situation" of the second system "must be independent of what happens" to the first



In discussions: Einstein and Bohr in 1930.

system. He decides that "One can escape this conclusion only by either assuming that measurement of one system (telepathically) changes the real situation of the other or by denying independent real situations to such separated systems". Both alternatives appear to him entirely unacceptable.

Bohr, in his essay, shows beautiful drawings of the *gedanken* experimental set-ups, like the famous two-slit and photon-in-a-box arrangements, suggested by Einstein in order to show that there is something wrong with

Similarly misleading are the anachronistic classification of soil moisture into distinct categories and the definitions of such concepts as 'field capacity' and 'wilting points' (the latter erroneously equated with hygroscopic moisture). In recent decades, these static concepts have been superseded by an understanding of the interactive dynamics between water and the soil-plant-atmosphere continuum.

The author's strong opinions occasionally interfere with the objective treatment of the subject; an example is her blanket condemnation of dam projects. And any book necessarily reflects its author's experience and perception. As Pielou is Canadian, it is natural that her book should include an extensive (and interesting) treatment of freezing and thawing phenomena. On the other hand, the book does not give a proportionate treatment of water relations in the semi-arid, arid and extremely arid (desert) regions that cover so much of our globe and in which so many nations struggle to subsist. Notably absent also is any description of tropical regions (such as the tropical rainforests), where the land,

A planet's lifeblood

Images of water in all its forms can be found in *Water: Worlds Between Heaven and Earth* (Stewart, Tabori & Chang, \$50, £35), with photographs by Art Wolfe and text by Michelle A. Gilders and Claus Biegert. Publication of the book is scheduled for 31 May.



water and biotic resources are particularly vulnerable to destruction.

Such criticisms notwithstanding, I found Pielou's book a welcome addition to the genre of literature designed to bridge the gap

between scientists, with their specialized jargon and occasionally obtuse theories, and the intelligent and concerned lay public. In a democratic society, communicating sound science to the public is essential if we are to enlist wide support for the actions necessary to resolve the environmental issues involving water that confront nations and the international community with increasing severity and urgency.

The World's Water by Peter H. Gleick apparently has the same aim as Pielou's book, though it differs greatly in structure and content. Unlike *Fresh Water*, and despite its seemingly inclusive title, it is not in itself a complete book. Rather, it is a somewhat disjointed collection of disparate items, including data gleaned from UN agencies and the World Bank, reviews of recent developments on water projects and their management, conjectured prospects for the future, and editorial opinions of the various topics covered.

Given Gleick's earlier seminal contribution to the field (*Water in Crisis*, Oxford University Press; 1993), we might have expected fuller coverage. However, the selection of topics and items is somewhat arbitrary. The exposition attempts to be authoritative, but is rather journalistic in tone and style. The whole appears to be an effort to emulate the annual WorldWatch "State of the World" publications addressed to a more specific topic.

What this volume does most laudably is argue the case for a shift of emphasis from the supply-side of water management to the demand-side, stressing the need and potential for improved efficiency of water use through conservation, recycling and environmental sustainability. Another important issue mentioned, if not fully resolved, is the treatment of water as an economic commodity. Pricing water may very well promote efficient use but might

NIELS BOHR ARCHIVE

quantum physics. In all these cases, Bohr was able to prove Einstein was wrong by analysing exactly what 'observation' means. He emphasizes how important it is in physics to analyse precisely what can be said about an experiment. Bohr is acutely aware of the huge potential for misunderstanding and renounces any allusions to mysticism. He also dislikes phrases like "disturbing phenomena by observations" because of their potential for confusion. Bohr stresses the use of "the word *phenomenon* exclusively to refer to observations made under specific circumstances, including an account of the whole experimental arrangement". Einstein's desideratum of an independent, real situation then becomes devoid of meaning for his separated quantum systems.

Einstein's philosophical beliefs are carefully analysed in the book by a number of eminent philosophers, for example Philipp Frank, a member of the Vienna Circle. And the essays by V. F. Lenzen and F. S. C. Northrop systematizing some of Einstein's views on the theory of knowledge and his conception of science even merit the master's approval in his "Reply" at the end of the collection. These essays, and Hans Reichenbach's analysis of the philosophical significance of the theory of relativity, make excellent reading for any philosophy seminar.

It is impossible to do justice to all the contributions. Most impressive are Kurt Gödel's proposal of closed-time-like solutions to general relativity which allow one to return from the past even when progressing all the time into the future, and Carl Menger's suggestion that we might ultimately have to give up the idea of a

continuum. Georges Lemaître's discussion of the cosmological constant is an example of a topic that has recently come into focus again because of observed hints of an anomalous expansion of the Universe. Lemaître wisely notes that "The history of science provides many instances of discoveries which have been made for reasons which are no longer considered satisfactory".

In that spirit, we note that quantum entanglement is one such case. Einstein clearly foresaw its enormous philosophical consequences, which motivated him to introduce it in his critique of quantum physics. While that critique can hardly be upheld today, entanglement became the central concept in the newly emerging field of quantum information, which includes such eye-catching topics as teleportation and quantum computation. It is tantalizing to contemplate Einstein's and Bohr's reactions to the enormous experimental progress made in recent years. Realization of many of the old *gedanken* experiments and their refinements have given rise to a generation of physicists to whom experiments with individual quanta are an everyday experience in the laboratory, and who have thus obtained a natural, intuitive understanding of quantum phenomena.

Bohr, in his essay, mentions the "benefit from the inspiration which we all derive from every contact with Einstein". This, I promise, will be true for every patient reader of the book, who has a wonderful opportunity to encounter one of the greatest minds of all time. □

Anton Zeilinger is at the Institute of Experimental Physics, Vienna University, Boltzmanngasse 5, 1090 Vienna, Austria.

not ensure that long-term environmental considerations are properly taken into account.

In a section entitled "A vision for 2050", Gleick presents an optimistic future, presumably conditional on the timely adoption of appropriate policies. This is a useful exercise, inasmuch as it offers hope rather than gloom. It would have been even more useful, however, had the author mapped out alternative scenarios (at least semi-quantitatively), comparing and contrasting the probable outcomes of specified circumstances (for example, rising or falling energy costs, or climatic changes resulting in increased or reduced availability of water) and policies (pricing of water and other incentives to promote water conservation). Gleick makes much of the concept of basic needs for water, but does not discuss how such needs are conditioned by climate, type of economy or mode of subsistence.

Perhaps the most useful section of the volume is the Data Section, where updated information is given in tabular form. However, in presenting the total renewable freshwater supply for each country, the data are seriously flawed in failing to subtract the volume of flow discharged to downstream countries. As a result, the same volumes of water are counted several times over (as in the case of Ethiopia, Sudan and Egypt, for example). As the data are taken from various sources, critical evaluations of their reliability would also have been helpful.

As this is only the first of what is evidently intended to constitute a series of books to be published every two years, we can only hope that subsequent volumes will provide a more complete coverage of the topics

and issues that would be properly included under the encompassing title *The World's Water*. □

Daniel Hillel is Professor Emeritus of Environmental Physics and Hydrology, University of Massachusetts, Amherst, Massachusetts 01003, USA.

Sounding out two centuries

Sounds of Our Times: Two Hundred Years of Acoustics

by Robert T. Beyer

Springer: 1999. 444 pp. \$49.95, £37.50

J. Woodhouse

The science of sound and vibration has come down in the world during the twentieth century. It occupied a position of pre-eminence in antiquity, with the Pythagorean views of universal harmony and the 'music of the spheres'.

By the nineteenth century it was still in the mainstream of rapidly developing physical science, and many famous names made contributions: scientists like Hermann von Helmholtz, Lord Rayleigh, Gustav Kirchhoff and Joseph Henry, and inventors like Alexander Graham Bell and Thomas Edison. In the twentieth century the subject has lacked the glamour of quantum mechanics, astrophysics or molecular biology, but nevertheless research has continued with accelerating pace.

Robert Beyer's engagingly written history of the subject since 1800 tells the whole story, and may help restore some sparkle to a Cinderella subject. It is liberally supplied

with period illustrations and anecdotes, while maintaining a high standard of technical accuracy and completeness. Among the many gems is the first experimental verification of the formula for the Doppler effect, when the experimental apparatus consisted of a steam locomotive, several trumpeters and some musically trained observers. Beyer also describes the long story of ways to visualize and measure waveforms of sound and vibration — from Helmholtz resonators, tuning-forks and rotating mirrors through the early developments in electromagnetic devices, up to modern electronic and computer-based methods.

During this century the scope of the subject has expanded enormously, with such topics as underwater acoustics, architectural acoustics and ultrasonics. All are covered in Beyer's systematic account, which draws especially on the author's long association with the Acoustical Society of America whose journal and regular conferences provide invaluable raw material, while the society itself is included as part of the history of the subject.

The book comes right up to date with an inevitably brief overview of recent developments and trends across the whole field, from structural vibration to the physiology of hearing. The result is of interest to any scientific reader. Those working in the subject will naturally turn first to the reference list to see if they are mentioned. I wasn't, but I was quite impressed at the choice of modern developments in the areas with which I am familiar.

It would be unrealistic to expect such a general book to give a definitive review of every speciality, but Beyer has made a very creditable stab. This book combines good science with being a good read. It should make its way into many libraries, and onto many bedside tables. □

J. Woodhouse is in the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK.

New in paperback

Hunting Down the Universe: The Missing Mass, Primordial Black Holes and Other Dark Matters

by Michael Hawkins with Celia Fitzgerald
Perseus, \$13

"...his hypothesis is very likely to be disproved to the satisfaction of any sapient individual. If, on the other hand, his theory is observationally confirmed, then he will without doubt be acclaimed by the establishment, and I will eat this review." William Press, *Nature* **388**, 138 (1997).

About Face

by Jonathan Cole

MIT Press, \$16.81, £15.85

"Cole believes that facial expressions are the main way to communicate emotion, and to prove his point he has had the ingenious idea of examining the deficits of those who either

cannot interpret the expressions of others ... or cannot display such expressions themselves." Stuart Sutherland, *Nature* **390**, 458 (1997).

The Universe Below: Discovering the Secrets of the Deep Sea

by William J. Broad, illustrations by Dimitry Schidlovsky
Simon & Schuster, \$15, £9.99

The Thermal Warriors: Strategies of Insect Survival
by Bernd Heinrich

Harvard University Press, \$17.95

Visions: How Science will Revolutionize the 21st Century and Beyond

by Michio Kaku

Oxford University Press, £8.99, \$14



An early sound-recording session.

Sublimation from icy jets as a probe of the interstellar volatile content of comets

Geoffrey A. Blake^{*†}, C. Qi^{*}, Michiel R. Hogerheijde[‡], M. A. Gurwell[§] & D. O. Muhleman^{*}

^{*} Division of Geological and Planetary Sciences, California Institute of Technology 150-21, Pasadena, California 91125, USA

[†] JILA, University of Colorado, Boulder, Colorado 80309, USA

[‡] Department of Astronomy, University of California, Berkeley, California 94720, USA

[§] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

Comets are some of the most primitive bodies left over from the Solar System's early history. They may preserve both interstellar material and material from the proto-solar nebula, and so studies of their volatile components can provide clues about the evolution of gases and ices, as a collapsing molecular cloud transforms into a mature planetary system^{1,2}. Previous observations of emission from rotational transitions in molecules have averaged over large areas of the inner coma, and therefore include both molecules that sublimated from the nucleus and those that result from subsequent chemical processes in the coma. Here we present high-resolution observations of emission from the molecules HNC, DCN and HDO associated with comet Hale-Bopp. Our data reveal arc-like structures—icy jets—offset from (but close to) the nucleus. The measured abundance ratios on 1–3'' scales are substantially different from those on larger scales^{3–5}, and cannot be accounted for by models of chemical processes in the coma^{2,6,7}; they are, however, similar to the values observed in the cores of dense interstellar clouds and young stellar objects. We therefore propose that sublimation from millimetre-sized icy grains ejected from the nucleus provides access to relatively unaltered volatiles. The D/H ratios inferred from our data suggest that, by mass, Hale-Bopp (and by inference the outer regions of the early solar nebula) consists of ≥ 15 –40% of largely unprocessed interstellar material.

HNC is a metastable isomer of HCN, and is found in negligible amounts in equilibrium models of the solar nebula. Its detection in comets is thus an important indicator of kinetically controlled processes, that is, chemistry which is driven by the rates of various reactions and not by thermodynamics. Ion–molecule networks in dense molecular clouds^{2,8,9} lead to HNC/HCN ratios of ≈ 0.2 –1, while in analogues of the solar nebula, the T Tauri star accretion disks, values near 0.4 are seen¹⁰. In certain chemical kinetic models of these disks, the high HNC abundances are produced by photochemistry, while in others they arise from surviving molecular cloud material¹¹. HNC can also be synthesized within the coma^{6,7}, and the radial dependence of the HNC/HCN ratio in comets should be a sensitive function of direct HNC outgassing versus *in situ* production¹².

D/H ratios can provide even more sensitive diagnostics of cometary origins. In dense clouds, trace species exhibit D/H enrichments of several orders of magnitude that are both molecule specific and temperature sensitive. The typical interstellar values are much larger than those expected for volatiles synthesized in protoplanetary regions of the solar nebula. For example, D/H ratios of 0.01–0.05 are common in molecular cloud organic compounds¹³, while values of up to 0.18 have been detected towards low-mass protostars⁹. HDO/H₂O ratios are difficult to measure in cold regions, but lower bounds are provided by the values seen in hot cores¹⁴ (2 – 6×10^{-4}). These minimum HDO enhancements are still ten times the estimated protosolar deuterium fraction¹⁵ of (D/H)_{proto} $\approx 3 \times 10^{-5}$, and are similar to the water D/H values

measured in comets Halley, Hyakutake and Hale-Bopp (3.1 , 2.9 , 3.3×10^{-4}) (refs 5, 16, 17). High-spatial-resolution studies of the molecular D/H variability should therefore provide clues as to the interstellar versus nebular contributions to cometary nuclei.

The HCN and HNC $1 \rightarrow 0$ transitions were observed simultaneously on 31 March 1997 with the Owens Valley Radio Observatory (OVRO) millimetre-array. Whereas the peak of the HCN emission is always found to lie within the width of a beam of the nucleus, weaker, more extended arcs are often observed whose position angles rotate with a period (P_{nucleus}) and phase that are

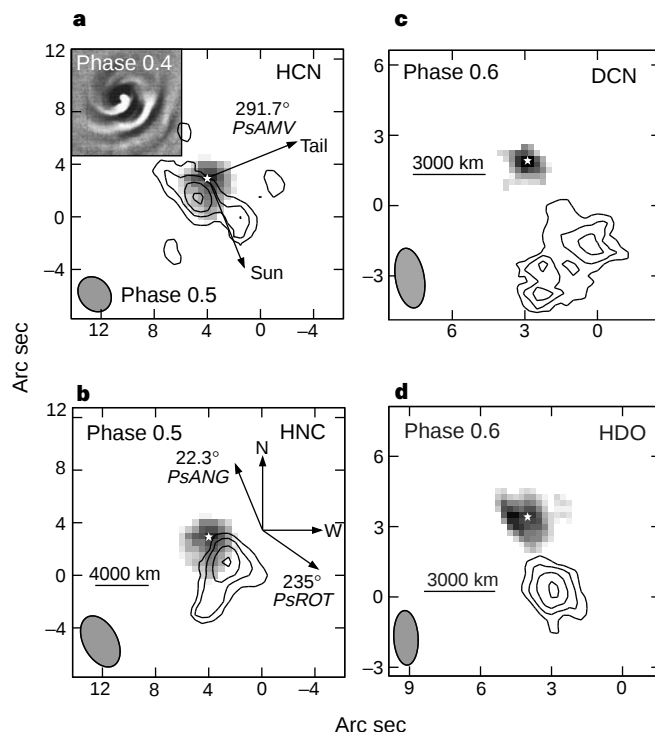


Figure 1 Contour images of millimetre-wave rotational line emission from the coma of comet Hale-Bopp. The full-width at half-maximum of the synthesized beam is shown by the ellipse in the lower left corner of each panel. Offsets (in arcsec) from the array pointing centre are indicated along the vertical and horizontal axes. Visibility phases and amplitudes were calibrated using interleaved observations of the quasar 0133+476, whereas measurements of Uranus, Neptune and the quasar 3C273 provided an absolute flux scale and determined the digital correlator passband. **a**, HCN $J = 1 \rightarrow 0$, $F = 2 \rightarrow 1$ (88.6318 GHz) emission. Contours start at $0.42 \text{ Jy beam}^{-1} \text{ km s}^{-1}$ and are spaced by $0.21 \text{ Jy beam}^{-1} \text{ km s}^{-1}$. Arrows indicate the sky position angles (counter-clockwise from north) for the solar direction and the minus velocity vector (PsAMV) of the nucleus, while the upper left corner contains an optical image of the Hale-Bopp dust jets at a rotational phase of 0.4 on 2 April 1997 (ref. 19). **b**, HNC $1 \rightarrow 0$ (90.6635 GHz) emission. Contours start at $0.28 \text{ Jy beam}^{-1} \text{ km s}^{-1}$ and are spaced by $0.14 \text{ Jy beam}^{-1} \text{ km s}^{-1}$. Arrows depict the cardinal directions, the position angle of the comet spin axis (N pole, PsROT), and the anti-solar direction (PsANG). Both **a** and **b** were measured between 31 March 1997 17:15 and 20:00 UT and integrated across the full velocity range (-1.3 to $+1.3 \text{ km s}^{-1}$) in which emission is detected at the 2σ level. **c**, Velocity integrated (0 – 2.0 km s^{-1}) DCN $J = 3 \rightarrow 2$ (217.238 GHz) emission between 30 March 1997 17:30 and 22:30 UT. Contours start at $0.9 \text{ Jy beam}^{-1} \text{ km s}^{-1}$ and are spaced by $0.45 \text{ Jy beam}^{-1} \text{ km s}^{-1}$. **d**, Velocity integrated (-2.0 to 0 km s^{-1}) HDO $2_{11} \rightarrow 2_{12}$ (241.561 GHz) emission between 29 March 1997 19:30 and 30 March 00:30 UT. Contours start at $0.8 \text{ Jy beam}^{-1} \text{ km s}^{-1}$ and are spaced by $0.4 \text{ Jy beam}^{-1} \text{ km s}^{-1}$. In each panel, the grey scale depicts the simultaneous observations of the dust and nucleus thermal continuum emission, the peak of which is marked by the star, while the definition of the rotational phase of the nucleus follows that of Licandro *et al.*¹⁸. The slight northeasterly shift of the nucleus position here and in Fig. 2 is due to uncertainties in the ephemeris used at the time of the observations.

well matched to that of the dust jets ($P_{\text{nucleus}} = 11.34 \pm 0.02$ h) (refs 18, 19). Analogous features are also observed in HNC. As Fig. 1a and b show, although extended HCN and HNC features are detected at position angles similar to those of the dust jets (position angle $\approx 230^\circ$ in Fig. 1), over certain time ranges they are strongly anti-correlated in a radial sense.

We have calculated abundances towards the nucleus and at the jet positions using both local thermodynamic equilibrium (LTE, with a rotational temperature T_{ROT} of 115 K; ref. 3) and statistical equilibrium (SE) Monte Carlo codes (see Table 1). In the latter case, the molecular excitation is determined including both collisions and radiation, and a Monte Carlo approach is used to solve the line radiative transfer (ref. 20, and M.R.H. and F. van der Tak, manuscript in preparation) within a one-dimensional Hauser coma model at a total water production rate of $10^{31} \text{ mol s}^{-1}$. The model is sampled at the observational visibility plane, or (u , v), spacings, and the abundances are varied until agreement with the interferometric data is obtained. Accordingly, the effects of the varying spatial resolutions at wavelengths of $\lambda = 3.3$ and 1.4 mm and the finite array sampling of extended structures are rigorously accounted for. Densities in the inner coma are sufficiently high that departures from LTE are found to be quite small^{3,5}, while the SE models show that all of the transitions studied here are optically thin.

For HCN, the observed hyperfine ratios of the $1 \rightarrow 0$ transition are also consistent with low optical depths, as is true in larger beams¹². Due to the large size of the coma and the coarse (u , v)-sampling of the present data sets, only 15–20% of the total flux within 45–60'' beams is detected. However, Hauser model fits with a single HCN production rate reproduce the OVRO $1 \rightarrow 0$, UMass 14 metre $1 \rightarrow 0$, and James Clerk Maxwell Telescope (JCMT) $4 \rightarrow 3$ intensities near perihelion^{6,12}, and demonstrate that the appropriate fluxes are recovered by the array even for sources which extend well beyond the primary beam of the six 10.4-m antennae. From images such as those in Fig. 1a we estimate that at most 20–25% of the total HCN $1 \rightarrow 0$ emission within 5'' of the nucleus arises in jet-like structures. When the extended coma emission and finite array (u , v)-sampling is accounted for, theoretically this fraction drops to $\sim 15\%$.

For the other species, the observed anisotropy (Fig. 1b–d) renders spherically symmetric derivations of the production rates inappropriate. To account for this anisotropy, estimates of the jet solid-angle filling factors are needed. Here, the column densities in Table 1 have been scaled by the ratio of the jet/nucleus linewidths (≈ 0.5 – 0.6), an admittedly simplistic approach which should result in upper bounds to the jet production rates. The calculated HNC abundances are smaller than those observed by single dishes, and lend support to the idea that some HNC must be synthesized in the extended coma^{6,7}. But the radial variation of a factor of 3 seen in the

HNC/HCN column-density ratios over a few thousand kilometres within the HNC arc are difficult to reconcile with the coma chemistry models¹².

Another illustration of the importance of jet activity comes from mapping the DCN and HDO emission. Figure 1c and d present images of DCN $3 \rightarrow 2$ and HDO $2_{11} \rightarrow 2_{12}$ transitions on 30 March and 29 March, respectively, at rotational phases (≈ 0.5 – 0.6) similar to those for HCN and HNC. DCN $3 \rightarrow 2$ images at two different rotational phases on 30 March are given in Fig. 2. The DCN and HDO position angles and the rotational sense of the DCN arcs are consistent with the known jetting properties of Hale–Bopp. Although we do not have simultaneous HCN and DCN observations, the HCN column densities on 31 March are in excellent agreement with UMass $1 \rightarrow 0$ single-dish maps on 30 March (A. J. Lowell, personal communications). Using the 31 March HCN fluxes at similar rotational phases, the DCN/HCN production-rate ratios in the icy jets vary by factors of at least two but are always quite large, $(\text{D}/\text{H})_{\text{HCN,jet}} \leq 0.025$, compared with the value of 2.3×10^{-3} found by JCMT observations of the DCN $5 \rightarrow 4$ transition⁴. Ratios with the 29 March HCN data are even larger, as are the direct ratios of the observed column densities.

Absorption by atmospheric water vapour makes direct measurements of the HDO/H₂O ratio impractical, but the OVRO fluxes and the large-scale water outgassing (from centimetre measurements of OH; ref. 3) of 10^{31} s^{-1} yield a lower bound of $(\text{D}/\text{H})_{\text{H}_2\text{O,jet}} \geq 7.5 \times 10^{-4}$. To estimate the jet D/H ratios we use the simultaneously observed HCN emission as a proxy. Provided the HCN/H₂O ratio in the jets is similar to the value of 0.002 observed in single-dish beams^{3,21}, the interferometric HDO/HCN production rates translate into a $(\text{D}/\text{H})_{\text{H}_2\text{O,jet}}$ estimate of 1.5 – 2.5×10^{-3} . This latter value is again nearly an order of magnitude larger than that found in the 10.5'' beam of the JCMT⁵, and over 20 times the D/H ratio of the Earth's oceans.

What processes could cause such significant differences between the volatile ratios on single-dish and interferometric scales? It is now clear that grains that are larger than about 1 mm are lifted by jet activity^{21,22}. Once exposed to the Sun, these grains warm rapidly and any volatiles are released. Single-dish examinations of Hale–Bopp average over distances of at least 8,000–10,000 km, whereas the resolution and spatial filtering of the array enables the emission associated with the dust/gas jets close to the nucleus to be selectively highlighted.

Gases subliming from the nucleus must pass through warm surface layers, while those released from grains in the coma pass through less material but are exposed to intense solar radiation. Studies of crystalline ice in the 140–210 K range have shown that a viscous water layer coexists ubiquitously²³. Little is at present known about the stability of HNC against isomerization to HCN as it diffuses through warm ice, water, and dust, but the rapid radial

Table 1 Observations and abundance ratios

Date	Species, location	I (K km s ⁻¹)	N_{LTE} ($10^{13} \text{ mol cm}^{-2}$)	Q_{LTE} ($10^{26} \text{ mol s}^{-1}$)	Q_{SE} ($10^{26} \text{ mol s}^{-1}$)	R_{LTE}	R_{SE}
29/03/97	HCN, nuc.	13.1	2.8	66	130	NA	NA
29/03/97	HCN, jet	8.5	18	42	57	NA	NA
29/03/97	HDO, nuc.	<3.0	<52	<140	<150	<1.2	<0.8
29/03/97	HDO, jet	10.7	170	150	240	1.6	2.2
30/03/97	DCN, nuc.	<2.5	<0.6	<0.7	<1.7	<0.006	<0.008
30/03/97	DCN, jet	11.5–14.4	2.7–3.4	1.9–2.2	3–4	≤ 0.024	≤ 0.036
31/03/97	HCN, nuc.	20.8	44	120	200	NA	NA
31/03/97	HCN, jet	15.9	34	92	110	NA	NA
31/03/97	HNC, nuc.	3.8	5.2	14	15	0.12	0.08
31/03/97	HNC, jet	11.7	16	22	23	0.24	0.21

Integrated intensities (I) are shown for the $J = 1 \rightarrow 0$ transitions of HCN (specifically, the $F = 2 \rightarrow 1$ component) and HNC, the $J = 3 \rightarrow 2$ transition of DCN, and the $2_{11} \rightarrow 2_{12}$ transition of HDO. Thermal equilibrium column densities (N_{LTE}) are reported, as are spherically symmetric production rates calculated with either a local thermodynamic equilibrium (Q_{LTE}) or statistical equilibrium (Q_{SE}) Hauser model. For collisional rates in the SE model, the 'interstellar' cross-sections for collisions with H₂ molecules²⁷ are scaled to reproduce the total 'cometary' cross-sections recommended by Crovisier²⁸ and Meier et al.⁴. Use of the H₂ cross-sections themselves (which are roughly an order of magnitude smaller) only changes the production rates by less than 10% as the levels observed are nearly in LTE. Abundance ratios with respect to HCN (R_{LTE} , R_{SE}) are calculated using the production rates derived from the LTE and Monte Carlo models. To account for the jet anisotropy, production rates at positions offset from the nucleus have been calculated according to $Q_{\text{jet}}(X) = Q_{\text{Hauser}} \times \Delta v_X / \Delta v_{\text{nucleus}}$, where Δv_X is the measured velocity width of species X at the jet position, and $\Delta v_{\text{nucleus}}$ is the measured velocity width of HCN emission at the nucleus. NA, not applicable.

variability of the HNC/HCN ratio suggests that some alteration occurs. Further, the acidity of HCN is sufficient to exchange D/H even in layers of water only a few micrometres thick under laboratory conditions. The D/H ratio of HCN, and by inference the D/H ratio of other trace species with labile hydrogens (NH_3 , CH_3OH , HCOOH , and so on), subliming from the nucleus of comets may therefore be controlled by that of the main constituent, water.

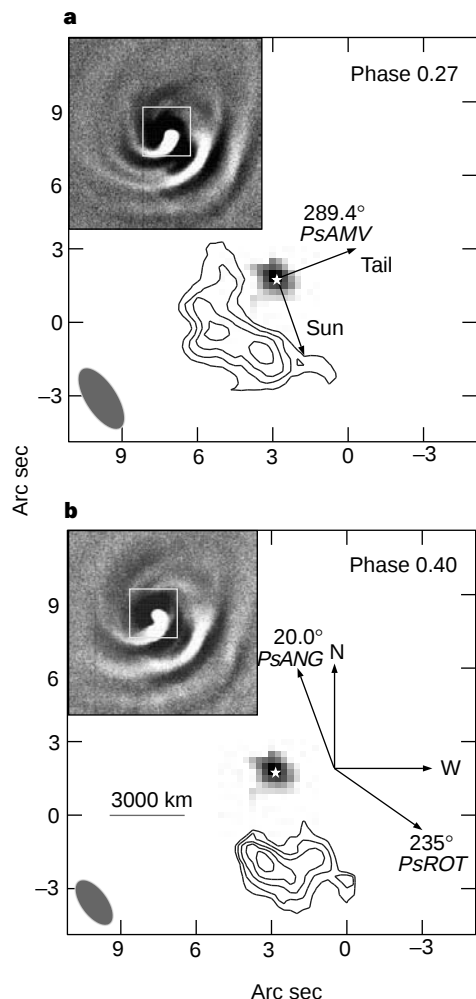


Figure 2 OVRO images of the $\text{DCN } J = 3 \rightarrow 2$ (217.238 GHz) emission in the coma of Hale-Bopp. Ellipses denote the full-width at half-maximum of the synthesized beam ($\approx 2.9'' \times 1.3''$), while the grey scale depicts the simultaneously measured dust continuum emission, the peak of which is denoted in the main panels by the star. Offsets from the array pointing centre are given in arc seconds. **a**, Velocity-integrated (-2.0 to 0 km s^{-1}) DCN emission between 30 March 1997 15:40 and 18:30 UT. The contours start at $0.66 \text{ Jy beam}^{-1} \text{ km s}^{-1}$ and are spaced by $0.33 \text{ Jy beam}^{-1} \text{ km s}^{-1}$. Arrows indicate the position angles of the solar direction and the minus velocity vector (PsAMV) of the comet. Inset, an optical image of dust jets in the coma of Hale-Bopp taken on 1 April 1997 at a rotational phase of 0.26 (ref. 19), with the box depicting the area shown in the OVRO image. **b**, Velocity-integrated (-2.0 to 0 km s^{-1}) DCN emission between 30 March 1997 16:30 and 20:30 UT. Contours start at $0.44 \text{ Jy beam}^{-1} \text{ km s}^{-1}$ and are spaced by $0.22 \text{ Jy beam}^{-1} \text{ km s}^{-1}$. Arrows depict the cardinal directions, the position angle of the comet's north pole (PsROT), and the anti-solar direction (PsANG). The inset shows a dust-jet optical image, taken on 2 April 1997 at a rotational phase of 0.4. For all the optical images, computer processing has been used to enhance the contrast and highlight the structure in the gas/dust distributions. While the DCN emission is entirely displaced from the nucleus, simultaneous observations of the $\text{H}_2\text{S } 2_{20} \rightarrow 2_{11}$ transition at 216.7104 GHz reveal intense emission centred on the dust continuum.

If such an exchange process occurs, those $(\text{D}/\text{H})_{\text{HCN}}$ ratios smaller than 0.025 that are measured in the DCN arcs may be more representative of the ices trapped within the nucleus; ices that would otherwise be detectable only with direct sampling techniques. As the region of high D/H gas is small and mixed with sublimation from the nucleus itself, it is difficult to detect in larger beams, suggesting that the D/H ratios in trace species with exchangeable hydrogen atoms should be viewed with caution when averaged over large spatial dimensions. To be consistent with the JCMT DCN/HCN ratio, the DCN jets can represent at most 10–15% of the total gas production rate. This is somewhat larger than the mass-loss rate of millimetre-sized grains from Hyakutake²², but is consistent with the HCN limits noted above.

Such large, differential D/H fractionations are difficult to produce under the conditions of high temperature and pressure found near jovian protoplanetary sub-nebulae². They are, however, entirely consistent with those seen in dense cloud cores at ~ 20 – 30 K . High, variable D/H ratios may also prevail in the outermost regions of the solar nebula where the chemical kinetic timescales can greatly exceed the dynamical lifetimes². Similar formation temperatures for individual species have been deduced for Hale-Bopp from measurements of the *ortho-para* ratio in water²⁴, and are consistent with the tentative detection of Ar at cosmogenic values²⁵.

Further estimates of the degree of mixing of protoplanetary and presolar material can be constrained by the observed $\text{HDO}/\text{H}_2\text{O}$ ratios. If the protosolar deuterium fraction¹⁵ is assumed to represent the (warm) nebular end member, a mass-balance calculation using the OVRO $\text{HDO}/\text{H}_2\text{O}$ upper and lower limits and the single-dish HDO measurements yields interstellar mass fractions of at least 15% and 40%. Higher nebular D/H values lower the instellar fraction, but would themselves be evidence of inherited material.

The high D/H ratios of the Hale-Bopp jets are therefore most simply explained by the survival of substantial reservoirs of presolar volatiles that did not equilibrate with the surrounding nebular gas. We stress that this does not imply that comets must contain assemblages of pristine interstellar ices²⁴. Sublimation of interstellar grain mantles as they pass through the nebular accretion shock is likely to occur, for example, as is radial mixing. Subsequent recondensation and grain growth is rapid, and the 50–70 K temperatures inferred from the abundances of volatiles such as CO, CH_4 and C_2H_6 are likely to measure the physical conditions under which the bulk of the nucleus was assembled²⁶, and need not be identical to molecular D/H or nuclear spin temperatures, which can be set at earlier times.

Received 28 October 1998; accepted 27 January 1999.

1. Festou, M. C., Rickman, H. & West, R. M. Comets. II. Models, evolution, origin, and outlook. *Astron. Astrophys. Rev.* **5**, 37–163 (1993).
2. Irvine, W. M., Schloerb, F., Crovisier, J., Fegley, B. & Mumma, M. in *Protostars & Planets IV* (eds Mannings, V., Boss, A. & Russell, S.) (Univ. Arizona Press, Tucson, in the press).
3. Biver, N. *et al.* Evolution of the outgassing of comet Hale-Bopp (C/1995 O1) from radio observations. *Science* **275**, 1915–1918 (1997).
4. Meier, R. *et al.* Deuterium in comet C/1995 O1 (Hale-Bopp): Detection of DCN. *Science* **279**, 1707–1710 (1998).
5. Meier, R. *et al.* A determination of the $\text{HDO}/\text{H}_2\text{O}$ ratio in comet C/1995 O1 (Hale-Bopp). *Science* **279**, 842–844 (1998).
6. Irvine, W. M. *et al.* Chemical processing in the coma as the source of cometary HNC. *Nature* **393**, 547–550 (1998).
7. Rodgers, S. & Charnley, S. HNC & HCN in comets. *Astrophys. J.* **501**, L227–250 (1998).
8. Schilke, P. *et al.* A study of HCN, HNC, and their isotopomers in OMC-1. *Astron. Astrophys.* **256**, 595–612 (1992).
9. Van Dishoeck, E. F., Blake, G. A., Jansen, D. J. & Groesbeck, T. D. Molecular abundances and low mass star formation. II. Organic and deuterated species toward IRAS 16293–2422. *Astrophys. J.* **447**, 760–782 (1995).
10. Dutrey, A., Guilloteau, S. & Guelin, M. Chemistry of protosolar-like nebulae: The molecular content of the DM Tau and GG Tau disks. *Astron. Astrophys.* **317**, L55–58 (1997).
11. Willacy, K., Klahr, H. H., Millar, T. J. & Henning, T. Gas and grain chemistry in a protoplanetary disk. *Astron. Astrophys.* **338**, 995–1006 (1998).
12. Irvine, W. M. *et al.* Chemistry in cometary comae. *Faraday Discuss.* **109**, 475–492 (1998).
13. Wootten, A. in *Astrochemistry* (eds Vardya, M. S. & Tarafdar, S. P.) 311–319 (IAU Symp. 120, Reidel, Dordrecht, 1987).
14. Jacq, T. *et al.* Deuterated water and ammonia in hot cores. *Astron. Astrophys.* **228**, 447–470 (1990).
15. Gautier, D. & Morel, P. A reestimate of the protosolar $(\text{D}/\text{H})_p$ ratio from $(^3\text{He}/^4\text{He})_\infty$ solar wind measurements. *Astron. Astrophys.* **323**, L9–L12 (1997).
16. Eberhardt, P., Reber, M., Krankowsky, D. & Hodges, R. R. The D/H and $^{18}\text{O}/^{16}\text{O}$ ratios in water from comet P/Halley. *Astron. Astrophys.* **302**, 301–316 (1995).

17. Bockelée-Morvan, D. *et al.* Deuterated water in comet C/1996 B2 (Hyakutake) and its implications for the origin of comets. *Icarus* **133**, 147–162 (1998).
18. Licandro, J. *et al.* The rotation period of C/1995 O1 (Hale-Bopp). *Astrophys. J.* **501**, L221–L225 (1998).
19. Mueller, B. E. A., Samarasinha, N. H. & Belton, M. J. S. Imaging of the structure and evolution of the coma morphology of Comet Hale-Bopp (C/1995 O1). *Earth, Moon, Planets* (in the press).
20. Hogerheide, M. R. The molecular environment of low-mass protostars. Thesis, Leiden Univ. (1998).
21. Bockelée-Morvan, D. & Rickman, A. C/1995 O1 (Hale-Bopp): Gas production curves and their interpretation. *Earth, Moon, Planets* (in the press).
22. Harmon, J. K. *et al.* Radar detection of the nucleus and coma of comet Hyakutake (C/1996 B2). *Science* **278**, 1921–1924 (1997).
23. Jenniskens, P., Banham, S. F., Blake, D. F. & McCoustra, M. R. S. Liquid water in the domain of cubic crystalline ice I_c. *J. Chem. Phys.* **107**, 1232–1241 (1997).
24. Crovisier, J. *et al.* The spectrum of comet Hale-Bopp (C/1995 O1) observed with the Infrared Space Observatory at 2.9 AU from the Sun. *Science* **275**, 1904–1907 (1997).
25. Stern, A., Slater, D. C., Festou, M. C., Parker, J. W. & A'Hearn, M. F. The EUV spectrum and the discovery of argon in comet Hale-Bopp. *Science* (submitted).
26. Notesco, G., Laufer, D. & Bar-Nun, A. The source of the high C₂H₂/CH₄ ratio in comet Hyakutake. *Icarus* **125**, 471–473 (1997).
27. Green, S., McLean, S. & McLean, A. D. Improved collisional excitation rates for interstellar water. *Astrophys. J. Suppl. Ser.* **85**, 181–185 (1993).
28. Crovisier, J. Rotational and vibrational synthetic spectra of linear parent molecules in comets. *Astron. Astrophys. Suppl. Ser.* **68**, 223–258 (1987).

Acknowledgements. We thank the staff at OVRO and D. K. Yeomans at JPL for support of our Hale-Bopp observations. The OVRO Millimeter Array is operated by the California Institute of Technology under funding from the US NSF. M.R.H. was supported by the Miller Institute for Basic Research in Science. G.A.B. was a Visiting Fellow at the University of Colorado; additional support to G.A.B. from the NASA Exobiology and Origins of Solar Systems programs is acknowledged.

Correspondence and requests for materials should be addressed to G.A.B. at the California Institute of Technology (e-mail: gab@gps.caltech.edu).

Gravity-wave interferometers as quantum-gravity detectors

Giovanni Amelino-Camelia*

Institut de Physique, Université de Neuchâtel, CH-2000 Neuchâtel, Switzerland

Nearly all theoretical approaches to the unification of quantum mechanics and gravity predict^{1–4} that, at very short distance scales, the classical picture of space-time breaks down, with space-time becoming somewhat ‘fuzzy’ (or ‘foamy’). The properties of this fuzziness and the length scale that characterizes its onset are potentially a means for determining which (if any) of the existing models of quantum gravity is correct. But it is generally believed⁵ that these quantum space-time effects are too small to be probed by technologies currently available. Here I argue that modern gravity-wave interferometers are sensitive enough to test certain space-time fuzziness models, because quantum space-time effects should provide an additional source of noise in the interferometers that can be tightly constrained experimentally. The noise levels recently achieved in one interferometer⁶ are sufficient to rule out values of the length scale that characterizes one of the space-time fuzziness models down to the Planck length ($\sim 10^{-35}$ m) and beyond, while the sensitivity required to test another model should be achievable with interferometers now under construction.

In a fuzzy space-time, the operative definition of a distance D is affected by quantum fluctuations. These fluctuations are primarily characterized by their overall magnitude σ_D (the root-mean-square deviation of D). The simplest proposals are such that:

$$\sigma_D \geq L_{\min} \quad (1)$$

where $L_{\min}$ is a quantum-gravity scale expected to be simply related to (usually identified with) the Planck length. Relations of the type shown in equation (1) are motivated by certain analyses of *gedanken* experiments (see, for example, ref. 7) that combine some elements of quantum mechanics and some elements of gravity. The quantum space-time fluctuations responsible for equation (1) are often

visualized as involving geometry and topology fluctuations¹, virtual black holes⁴ and other unusual phenomena.

Other models for space-time fuzziness arise when taking into account the quantum properties of devices, which were ignored in the original studies⁷ that led to the proposal of equation (1). It is well understood (see, for example, refs 8–13) that the combination of the gravitational properties and the quantum properties of devices can be important in the analysis of the operative definition of gravitational observables. The implications can be far-reaching; in particular, it was observed¹¹ that the masses of the probes used in measurements induce a change in the space-time metric, and this is associated with the emergence of non-locality. The nature of the gravitationally induced non-locality suggests¹¹ a modification of the fundamental commutators.

Here I shall be primarily concerned with the role that the quantum and the gravitational properties of devices have in the analysis of the measurability of distances, in the sense first illustrated by Wigner¹⁴. Wigner derived a quantum limit on the measurability of the distance D separating two bodies by analysing a measurement procedure based on the exchange of a light signal between the bodies. Taking into account Heisenberg’s position–momentum uncertainty relations for the clock used in the measurement procedure, Wigner obtained a lower bound on the quantum uncertainty in D :

$$\delta D \geq \sqrt{\frac{\hbar T_{\text{obs}}}{2M_c}} \approx \sqrt{\frac{\hbar D}{cM_c}} \quad (2)$$

where M_c is the mass of the clock, T_{obs} is the time required by the measurement procedure, and on the right-hand side I use the fact that the Wigner measurement of a distance D requires a time $2D/c$, where c is the speed of light.

The result shown in equation (2) may at first appear puzzling, as ordinary quantum mechanics should not limit the measurability of any given observable. (It only limits the combined measurability of pairs of conjugate observables.) However, quantum mechanics is the theoretical framework for the description of the outcome of experiments performed by classical devices. In the limit in which the devices (for example, Wigner’s clock) behave ‘classically’, which in particular requires the devices to be infinitely massive (so that $\delta x \delta v \approx \hbar/m \approx 0$), the right-hand side of equation (2) tends to zero. (Here δx and δv denote, respectively, the quantum uncertainties in the position and in the velocity of the device.) Therefore, as expected, there is no limitation on the measurability of the distance D in the appropriate infinite-mass ‘classical-device limit’. This line of argument depends crucially on the fact that ordinary quantum mechanics does not involve gravity.

The classical infinite-mass limit is clearly not consistent with the nature of measurements involving gravitational effects. As the devices get more and more massive they increasingly disturb the gravitational/geometrical observables, and well before reaching the infinite-mass limit the procedures for the measurement of gravitational observables cannot be meaningfully performed^{10,12,13}. A well-known example of this problem has been encountered in attempts (see, for example, ref. 8) to generalize to the study of the measurability of gravitational fields the Bohr–Rosenfeld analysis¹⁵ of the measurability of the electromagnetic field. To achieve the accuracy allowed by the formalism of ordinary quantum mechanics, the Bohr–Rosenfeld measurement procedure resorts to ideal test particles of infinite mass, which would of course not be admissible probes in a gravitational context. Similarly, in Wigner’s measurement procedure, the limit $M_c \rightarrow \infty$ is not admissible when gravitational interactions are taken into account. At the very least, the value of M_c is limited by the requirement that the clock should not turn into a black hole (which would not allow the required exchange of signals between the clock and the other devices). These observations, which render unavoidable the dependence on T_{obs} of equation (2), provide a basis for the possibility^{12,13} that in quantum

* Present address: Theory Division, CERN, CH-1211 Geneva, Switzerland. Permanent address: Dipartimento di Fisica, Università di Roma “La Sapienza”, Piazzale Moro 2, Roma, Italy.

gravity any measurement that monitors a distance D for a time T_{obs} is affected by quantum fluctuations characterized by:

$$\sigma_D \approx \sqrt{L_{\text{QG}} c T_{\text{obs}}} \quad (3)$$

where L_{QG} is a fundamental length scale which we can expect to be simply related to the Planck length. In particular, the Wigner measurement of a distance D , which requires a time $2D/c$, would be affected by fluctuations of magnitude $(L_{\text{QG}} D)^{1/2}$. (I note that the study reported in ref. 9 analysed the interplay of quantum mechanics and gravity in defining a net of time-like geodesics, and suggested that the maximum 'tightness' achievable for the geodesics net is $(L_{\text{QG}} D)^{1/2}$.)

A σ_D that increases with T_{obs} is not surprising for space-time fuzziness models; in fact, the same phenomena that would lead to fuzziness are also expected to induce "information loss"²⁴ (the information stored in a quantum system degrades as T_{obs} increases). The argument based on the Wigner set-up provides motivation to explore the specific form $\sigma_D \approx T_{\text{obs}}^{1/2}$ of this T_{obs} -dependence.

From the type of T_{obs} -dependence of equation (3), and the stochastic properties of the processes¹⁻⁴ expected to characterize a fuzzy space-time, it follows that the quantum fluctuations responsible for equation (3) should have amplitude spectral density $S(f)$ with the f^{-1} dependence typical of "random walk noise"¹⁶:

$$S(f) = f^{-1} \sqrt{L_{\text{QG}} c} \quad (4)$$

If indeed $L_{\text{QG}} \approx 10^{-35}$ m, from equation (4) one obtains $S(f) \approx f^{-1} (5 \times 10^{-14} \text{ m Hz}^{1/2})$. (One expects that this formula for the quantum-gravity-induced $S(f)$ could only apply to frequencies f significantly smaller than the Planck frequency c/L_P and significantly larger than the inverse of the timescale over which, even ignoring the gravitational field generated by the devices, the classical geometry of the space-time region where the experiment is performed manifests significant curvature effects.)

Before commenting on how the proposal of equations (3) and (4) compares with data from modern gravity-wave interferometers, I consider another space-time fuzziness scheme, which involves fluctuations of significantly different magnitude. This alternative scheme^{10,17} is essentially based on the observation that the uncertainty described by Wigner's equation (2) can be combined with a classical-gravity estimate of the uncertainty in the measurement of the distance D that results from the distortion of geometry associated with the gravitational field generated by the clock. While the uncertainty (equation (2)) decreases with M_c , the uncertainty induced by gravity increases with M_c , and combining the two uncertainties one finds a minimum total uncertainty of the type $\delta D \approx (\tilde{L}_{\text{QG}}^2 c T_{\text{obs}})^{1/3}$, where $\tilde{L}_{\text{QG}}$ is a length scale analogous to L_{QG} . As in the other analyses of the dependence of the Wigner measurement on the time of observation, one is then led to consider a fuzzy space-time with a corresponding dependence on T_{obs} :

$$\sigma_D \approx (\tilde{L}_{\text{QG}}^2 c T_{\text{obs}})^{1/3} \quad (5)$$

The associated amplitude spectral density is:

$$\tilde{S}(f) = f^{-5/6} (\tilde{L}_{\text{QG}}^2 c)^{1/3} \quad (6)$$

which, for $\tilde{L}_{\text{QG}} \approx 10^{-35}$ m, gives $\tilde{S}(f) = f^{-5/6} (3 \times 10^{-21} \text{ m Hz}^{1/3})$.

Each of the proposals in equations (1), (3) and (5) was obtained within a corresponding scheme for the interplay between quantum mechanics and weak-field gravity. This is an approach that has already proven successful in quantum-gravity research; in fact, the phenomenon of gravitationally induced phases, which was also predicted from analyses of the interplay between quantum mechanics and weak-field gravity, has already been confirmed experimentally¹⁸. By discovering experimentally which of the

space-time fuzziness proposals is correct, we could also obtain additional insight in the weak-field limit of quantum gravity, and the requirement of consistency with the correct weak-field limit can represent a highly non-trivial constraint for the search for quantum gravity. In particular, the popular quantum-gravity theories based on critical strings appear to require fuzziness of the type shown in equation (1), as seen in analyses of string collisions at planckian energies¹⁹; a proposal for a quantum-group structure which might accommodate equation (1) has been discussed in ref. 20. Equation (3) has been found to arise within the mathematical framework of dimensionally deformed Poincaré symmetries^{21,22}, which has been attracting interest recently. Point-particle quantum-gravity theories based on these deformations would therefore require space-time fuzziness of the type shown in equation (3). Moreover, equation (3) has been shown to hold within Liouville (non-critical) string theory^{3,23}, another approach to quantum gravity which is attracting interest. The search of quantum-gravity theories whose weak-field limit is consistent with equation (5) has not yet been successful, but there appears to be no obstruction in principle, so one can expect progress in this direction to be forthcoming.

Whereas conceptually the proposals shown in equations (1), (3) and (5) represent drastic departures from conventional physics, phenomenologically they appear to encode only minute effects; for example, it has been observed that, assuming L_{min} , L_{QG} and $\tilde{L}_{\text{QG}}$ are not much larger than the Planck length, all of these proposals encode fluctuations of <1 metre on the size of the whole observable Universe ($\sim 10^{10}$ light years). However, the precision²⁴ of modern gravity-wave interferometers is such that they could provide significant information (at least) on proposals (3) and (5). In fact, the operation of gravity-wave interferometers is based on the detection of minute changes in the positions of some test masses (relative to the position of a beam splitter). If these positions were affected by quantum fluctuations of the type discussed above, the operation of gravity-wave interferometers would effectively involve an additional source of noise due to quantum gravity. This observation already allows us to set bounds using existing noise-level data obtained at the California Institute of Technology 40-m interferometer. This interferometer has achieved⁶ displacement noise levels with amplitude spectral density lower than $10^{-18} \text{ m Hz}^{-1/2}$ for frequencies between 200 and 2,000 Hz and this, as seen by comparison with equation (4), rules out all values of L_{QG} down to the Planck length. In fact, even values of L_{QG} significantly lower than the Planck length are inconsistent with the data reported in ref. 6; in particular, by comparing equation (4) with the observed noise level of $3 \times 10^{-19} \text{ m Hz}^{-1/2}$ near 450 Hz, which is the best achieved at the above interferometer, one obtains the bound $L_{\text{QG}} \leq 10^{-40}$ m. Although at present we should allow for some relatively small factor to intervene in the relation between L_{QG} and L_P , having excluded all values of L_{QG} down to 10^{-40} m, the status of proposal (3) appears to be at best problematic. Even more-stringent bounds on L_{QG} are within reach of the next LIGO/VIRGO^{25,26} generation of gravity-wave interferometers.

The sensitivity achieved at the Caltech 40-m interferometer also sets a bound on proposal (5) and (6). By observing that equation (6) would imply quantum-gravity noise levels for gravity-wave interferometers of the order of $\tilde{L}_{\text{QG}}^{2/3} (10 \text{ m}^{1/3} \text{ Hz}^{-1/2})$ at frequencies of a few hundred hertz, one obtains from the data reported in ref. 6 that $\tilde{L}_{\text{QG}} \leq 10^{-29}$ m. This bound is remarkably stringent in absolute terms, but is still well above the range of values of $\tilde{L}_{\text{QG}}$ that is emerging from the various theoretical approaches to quantum gravity. A more significant bound on $\tilde{L}_{\text{QG}}$ should be obtained by the LIGO/VIRGO generation of gravity-wave interferometers. For example, it is plausible²⁵ that the 'advanced phase' of LIGO will achieve a displacement noise spectrum of less than $10^{-20} \text{ m Hz}^{-1/2}$ near 100 Hz, and this would probe values of $\tilde{L}_{\text{QG}}$ as small as 10^{-34} m.

Beyond the LIGO/VIRGO generation of gravity-wave interferometers, there are still quite sizeable margins for improvement by

optimizing the performance of the interferometers at low frequencies, where both equations (4) and (6) become more significant. It appears natural to perform such studies in the quiet environment of space, perhaps through future refinements of LISA-type set-ups²⁷.

The above discussion of gravity-wave interferometers shows that the smallness of the Planck length does not preclude the possibility of direct investigations of space-time fuzziness. This complements the results of studies^{28,29} which have shown that indirect evidence of quantum space-time fluctuations could be obtained by testing the predictions of theories consistent with a given picture of these fluctuations. Additional encouragement for experiment-driven progress in understanding the interplay between gravity and quantum mechanics comes from recent studies^{30,31} in the area of gravitationally induced phases, the significance of which has been emphasized in refs 32 and 33. □

Received 12 August; accepted 21 December 1998.

1. Wheeler, J. A. *Relativity, Groups and Topology* (eds De Witt, B. S. & De Witt, C. M.) (Gordon & Breach, New York, 1963).
2. Ashtekar, A., Rovelli, C. & Smolin, L. Weaving a classical geometry with quantum threads. *Phys. Rev. Lett.* **69**, 237–240 (1992).
3. Ellis, J., Mavromatos, N. & Nanopoulos, D. V. String theory modifies quantum mechanics. *Phys. Lett. B* **293**, 37–48 (1992).
4. Hawking, S. W., Page, D. N. & Pope, C. N. Quantum gravitational bubbles. *Nucl. Phys. B* **170**, 283–306 (1980).
5. Isham, C. in *Proc. 14th Int. Conf. on General Relativity and Gravitation* (eds Francavaglia, M., Longhi, G., Lusanna, L. & Sorace, E.) (World Scientific, Singapore, 1997).
6. Abramovici, A. *et al.* Improved sensitivity in a gravitational wave interferometer and implications for LIGO. *Phys. Lett. A* **218**, 157–163 (1996).
7. Garay, L. J. Quantum gravity and minimum length. *Int. J. Mod. Phys. A* **10**, 145–166 (1995).
8. Bergmann, P. G. & Smith, G. J. Measurability analysis of the linearized gravitational field. *Gen. Rel. Grav.* **14**, 1131–1166 (1982).
9. Diosi, L. & Lukacs, B. On the minimum uncertainty of space-time geodesics. *Phys. Lett. A* **142**, 331–334 (1989).
10. Ng, Y. J. & Van Dam, H. Limit to space-time measurement. *Mod. Phys. Lett. A* **9**, 335–348 (1994).
11. Ahluwalia, D. V. Quantum measurement, gravitation, and locality. *Phys. Lett. B* **339**, 301–303 (1994).
12. Amelino-Camelia, G. Limits on the measurability of space-time distances in the semi-classical approximation of quantum gravity. *Mod. Phys. Lett. A* **9**, 3415–3422 (1994).
13. Amelino-Camelia, G. Dimensionful deformations of Poincaré symmetries for a Quantum Gravity without ideal observers. *Mod. Phys. Lett. A* **13**, 1319–1325 (1998).
14. Wigner, E. P. Relativistic invariance and quantum phenomena. *Rev. Mod. Phys.* **29**, 255–268 (1957).
15. Bohr, N. & Rosenfeld, L. Zur Frage der Messbarkeit der Elektromagnetischen Feldgrößen. *Kongelige Danske Videnskabernes Selskab Matematisk-Fysiske Meddelelser* **12**, 1–65 (1933).
16. Radeka, V. Low-noise techniques in detectors. *Annu. Rev. Nucl. Part. Sci.* **38**, 217–277 (1988).
17. Karolyhazy, F. Gravitation and Quantum Mechanics of macroscopic objects. *Il Nuovo Cimento A* **42**, 390–402 (1966).
18. Colella, R., Overhauser, A. W. & Werner, S. A. Observation of gravitationally induced quantum interference. *Phys. Rev. Lett.* **34**, 1472–1474 (1975).
19. Amati, D., Ciafaloni, M. & Veneziano, G. Superstring collisions at Planckian energies. *Phys. Lett. B* **197**, 81–88 (1987).
20. Kempf, A., Mangano, G. & Mann, R. B. Hilbert space representation of the minimal length uncertainty relation. *Phys. Rev. D* **52**, 1108–1118 (1995).
21. Lukierski, J., Nowicki, A. & Ruegg, H. Classical and quantum-mechanics of free κ -relativistic systems. *Annu. Phys.* **243**, 90–116 (1995).
22. Amelino-Camelia, G. Enlarged bound on the measurability of distances and quantum κ -Poincaré group. *Phys. Lett. B* **392**, 283–286 (1997).
23. Amelino-Camelia, G., Ellis, J., Mavromatos, N. E. & Nanopoulos, D. V. Distance measurement and wave dispersion in a Liouville string approach to quantum gravity. *Int. J. Mod. Phys. A* **12**, 607–623 (1997).
24. Saulson, P. R. *Fundamentals of Interferometric Gravitational Wave Detectors* (World Scientific, Singapore, 1994).
25. Abramovici, A. *et al.* LIGO: The Laser Interferometer Gravitational-Wave Observatory. *Science* **256**, 325–333 (1992).
26. Bradaschia, C. *et al.* The VIRGO project: a wide band antenna for gravitational wave detection. *Nucl. Instrum. Meth. A* **289**, 518–525 (1990).
27. Danzmann, K. LISA: Laser interferometer space antenna for gravitational wave measurements. *Class. Quant. Grav.* **13**, A247–A250 (1996).
28. Ellis, J., Hagelin, J. S., Nanopoulos, D. V. & Srednicki, M. Search for violations of Quantum Mechanics. *Nucl. Phys. B* **241**, 381–405 (1984).
29. Amelino-Camelia, G., Ellis, J., Mavromatos, N. E., Nanopoulos, D. V. & Sarkar, S. Tests of quantum gravity from observations of γ -ray bursts. *Nature* **393**, 763–765 (1998).
30. Ahluwalia, D. V. Can general relativistic description of gravitation be considered complete? *Mod. Phys. Lett. A* **13**, 1393–1400 (1998).
31. Snadden, M. J., McGuirk, J. M., Bouyer, P., Haritos, K. G. & Kasevich, M. A. Measurement of the Earth's gravity gradient with an atom interferometer-based gravity gradiometer. *Phys. Rev. Lett.* **81**, 971–974 (1998).
32. Service, R. F. Gravity measurements ride the atom wave. *Science* **281**, 762–763 (1998).
33. Seife, C. Einstein in free fall. *New Sci.* **158**, 111 (1998).

Acknowledgements. I thank A. Ashtekar for suggesting that gravity-wave interferometers might be useful for experimental tests of some quantum-gravity phenomena; D. Ahluwalia, J. Ellis, J. Lukierski, N. E. Mavromatos, C. Rovelli, S. Sarkar and J. Stachel for discussions about quantum-gravity models; and F. Barone, M. Coles, J. Faist, R. Flaminio, L. Gammaiton, G. Gonzalez, T. Huffman, L. Marrucci and M. Punturo for conversations about experimental interferometry. This work was supported by the Swiss National Science Foundation.

Correspondence and requests for materials should be addressed to the author (e-mail: giovanni.amelino-camelia@cern.ch).

Four-wave mixing with matter waves

L. Deng^{††}, E. W. Hagley^{*}, J. Wen^{*}, M. Trippenbach^{*‡}, Y. Band^{*‡}, P. S. Julienne^{*}, J. E. Simsarian^{*}, K. Helmerson^{*}, S. L. Rolston^{*} & W. D. Phillips^{*}

^{*} Atomic Physics Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

[†] Department of Physics, Georgia Southern University, Statesboro, Georgia 30460, USA

[‡] Department of Chemistry and Physics, Ben-Gurion University of the Negev, Beer-sheva 84105, Israel

The advent of the laser as an intense source of coherent light gave rise to nonlinear optics, which now plays an important role in many areas of science and technology. One of the first applications of nonlinear optics was the multi-wave mixing^{1,2} of several optical fields in a nonlinear medium (one in which the refractive index depends on the intensity of the field) to produce coherent light of a new frequency. The recent experimental realization of the matter-wave 'laser'^{3,4}—based on the extraction of coherent atoms from a Bose–Einstein condensate⁵—opens the way for analogous experiments with intense sources of matter waves: nonlinear atom optics⁶. Here we report coherent four-wave mixing in which three sodium matter waves of differing momenta mix to produce, by means of nonlinear atom–atom interactions, a fourth wave with new momentum. We find a clear signature of a four-wave mixing process in the dependence of the generated matter wave on the densities of the input waves. Our results may ultimately facilitate the production and investigation of quantum correlations between matter waves.

The analogy between nonlinear optics with lasers and nonlinear atom optics with Bose–Einstein condensates can be seen in the similarities between the equations that govern each system. For a condensate of interacting bosons, in a trapping potential V , the macroscopic wavefunction Ψ satisfies a nonlinear Schrödinger equation⁷

$$i\hbar \frac{\partial \Psi}{\partial t} = \left(-\frac{\hbar^2}{2M} \nabla^2 + V + U_0 |\Psi|^2 \right) \Psi \quad (1)$$

where M is the atomic mass, U_0 describes the strength of the atom–atom interaction ($U_0 > 0$ for sodium atoms), and $|\Psi|^2$ is proportional to atomic number density. The nonlinear term $U_0 |\Psi|^2 \Psi$ in equation (1) is similar to the third order term $\chi^{(3)} |E|^2 E$ in the wave equation for the electric field E describing optical four-wave mixing (4WM; where the susceptibility $\chi^{(3)}$ depends on the nonlinear medium). We therefore expected 4WM with coherent matter waves, analogous to optical 4WM. In contrast to optical 4WM, the nonlinearity in matter-wave 4WM comes from atom–atom interactions; there is no need for an additional nonlinear medium.

The first theoretical study of nonlinear atom optics was reported in 1993⁸, and the idea of 4WM using condensates prepared in different electronic states to enhance the nonlinearity was discussed in 1995⁹. A recent calculation⁹ showed that the nonlinearity associated with the interaction between ground-state atoms is large enough to observe 4WM with wavepackets created from existing Bose–Einstein condensates. To produce matter-wave mixing, we create three overlapping wavepackets with momenta $\mathbf{P}_n$ ($n = 1, 2, 3$) and observe the creation of the 4WM wavepacket $\mathbf{P}_4$ that satisfies energy, momentum and particle-number conservation (Fig. 1).

In our experiment, we use Bragg diffraction of atoms from a moving optical standing wave¹⁰ to create the necessary three wavepackets, starting from a Bose–Einstein condensate. Briefly, we first form a condensate of $\sim 2 \times 10^6$ sodium atoms in the

$3S_{1/2}$, $F = 1$, $m = -1$ state using a combination of laser cooling and radio-frequency-induced evaporative cooling in a TOP (time-orbiting-potential) trap¹¹, without a discernible non-condensed fraction. We then adiabatically expand the potential¹⁰ in 4 s by simultaneously reducing the magnetic field gradient and increasing the rotating bias field. This reduces the trap frequencies in the $\hat{x}$, $\hat{y}$ and $\hat{z}$ directions to 84, 59 and 42 Hz, respectively. The asymptotic r.m.s. momentum width of the released condensate after adiabatic expansion is measured to be $0.14(\pm 0.02)\hbar k$ (all uncertainties reported here are one standard deviation combined statistical and systematic uncertainties). Because this is small compared to $\sqrt{2}\hbar k$, the smallest momentum imparted to the condensate with the Bragg diffraction, the wavepackets will spatially separate as the system evolves.

After adiabatic expansion, we switch off the trap, wait $600\ \mu\text{s}$ so that the trapping magnetic fields decay away and then apply a sequence of two Bragg pulses. Each $30\text{-}\mu\text{s}$ pulse is composed of two linearly polarized laser beams detuned from the $3S_{1/2}$, $F = 1 \rightarrow 3P_{3/2}$, $F' = 2$ transition by $\Delta/2\pi = -2\text{ GHz}$ to suppress spontaneous emission. This large detuning makes negligible Bragg scattering of the optical waves by the atoms, which could lead to a spurious scattering of atoms into P_4 . The frequency difference

between the two laser beams of a single Bragg pulse is chosen to fulfil a first-order Bragg diffraction condition that changes the momentum state of the atoms without changing their internal state¹⁰. The first Bragg pulse is composed of two mutually perpendicular laser beams of frequencies ν_1 and $\nu_2 = \nu_1 - 50\text{ kHz}$, and wavevectors $\mathbf{k}_1 = k\hat{x}$ and $\mathbf{k}_2 = -k\hat{y}$ ($k = 2\pi/\lambda$, $\lambda = 589\text{ nm}$). The maximum intensity of each beam is $\sim 10\text{ mW cm}^{-2}$. The intensity was chosen so that roughly 1/3 of the condensate atoms acquire momentum $\mathbf{P}_2 = \hbar(\mathbf{k}_1 - \mathbf{k}_2) = \hbar k(\hat{x} + \hat{y})$. The second Bragg pulse is applied $20\ \mu\text{s}$ after the end of the first Bragg pulse (well before the wavepackets are separated). This second Bragg pulse is composed of two counter-propagating laser beams with frequencies ν_1 and $\nu_3 = \nu_1 - 100\text{ kHz}$, and wavevectors $\mathbf{k}_1 = k\hat{x}$ and $\mathbf{k}_3 = -k\hat{x}$. The intensities of these laser beams were chosen to cause half of the remaining atoms in the momentum state $\mathbf{P}_1 = 0$ to acquire a momentum $\mathbf{P}_3 = \hbar(\mathbf{k}_1 - \mathbf{k}_3) = 2\hbar k\hat{x}$ (atoms in $\mathbf{P}_2$ are not affected by this pulse because of the Doppler shift of the light). We chose this pulse sequence so that only $\mathbf{P}_2 = \hbar k(\hat{x} + \hat{y})$ and $\mathbf{P}_3 = 2\hbar k\hat{x}$ are produced from $\mathbf{P}_1 = 0$. Thus we create, nearly simultaneously, three overlapping wavepackets of the requisite momenta. Without the nonlinear term in equation (1), one would expect only to observe these three wavepackets after they have spatially separated. But as

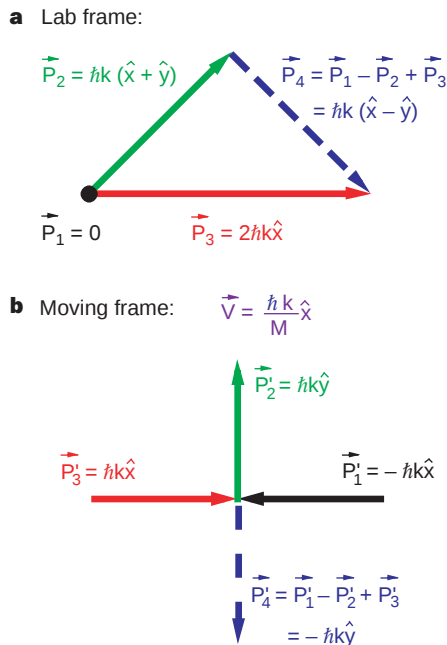


Figure 1 Momentum-energy conservation for 4WM and the bosonic stimulation viewpoint in a moving frame. **a**, Momentum conservation, $\mathbf{P}_4 = \mathbf{P}_1 - \mathbf{P}_2 + \mathbf{P}_3$ (equivalent to phase-matching in optical 4WM), in the laboratory frame. For clarity, over-arrows indicate vectors. Energy conservation requires $\mathbf{P}'_4 = \mathbf{P}'_1 - \mathbf{P}'_2 + \mathbf{P}'_3$. **b**, It is always possible to view matter 4WM in a frame moving with velocity $\mathbf{v}$ such that the three input momenta have the same magnitude, and two are counter-propagating. Then, in our case two atoms in momentum states $\mathbf{P}'_1 = -\hbar k\hat{x}$ and $\mathbf{P}'_3 = \hbar k\hat{x}$ are bosonically stimulated by wavepacket $\mathbf{P}'_2 = -\hbar k\hat{y}$ to scatter into momentum states $\mathbf{P}'_2$ and $\mathbf{P}'_4 = -\mathbf{P}'_2 = \hbar k\hat{y}$. We note that the energy and momentum conditions are satisfied independent of the direction of $\mathbf{P}'_2$. The 4WM wavepacket is a consequence of energy, momentum and particle-number conservation when atoms are stimulated into the momentum state $\mathbf{P}'_2$. Thus 4WM can be viewed as the annihilation of momentum states $\mathbf{P}'_1$ and $\mathbf{P}'_3$, and the creation of momentum states $\mathbf{P}'_2$ and $\mathbf{P}'_4$ (the minus signs in the energy and momentum conditions are attached to the state that gains atoms). It is this bosonic stimulation of scattering that mimics the stimulated emission of photons from an optical nonlinear medium. Alternatively, by choosing a frame of reference in which $\mathbf{P}'_1 = -\mathbf{P}'_2$ (or $\mathbf{P}'_2 = -\mathbf{P}'_3$), 4WM can also be viewed as matter-wave Bragg diffraction of $\mathbf{P}'_3$ ($\mathbf{P}'_1$) from the grating produced by the interference of two others.

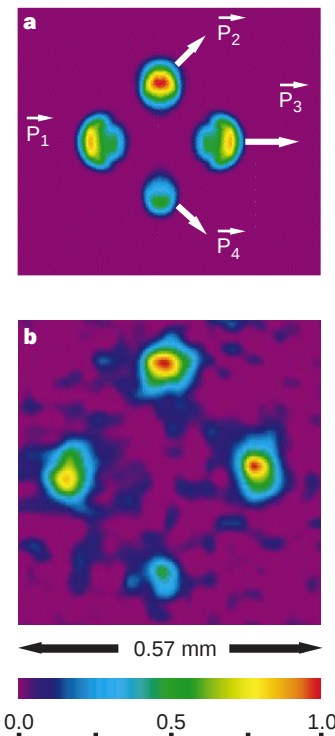


Figure 2 Numerical simulation and experimental results for 4WM. **a**, Calculated two-dimensional atomic distribution after 1.8 ms, showing the 4WM. The calculations were performed only until the wavepackets completely separated due to constraints on the simulation grid-size. The momenta are those of Fig. 1a. The field of view is $0.23 \times 0.26\text{ mm}$. We note that atoms are removed primarily from the back-end of the wavepackets because these regions overlap for the longest time. **b**, A false-colour image of the experimental atomic distribution showing the four (small) wavepacket generated by the 4WM process. The four wavepackets form a square measuring $0.26 \times 0.26\text{ mm}$, corresponding to the distance of 0.25 mm calculated using the experimental time of flight of 6.1 ms and the wavepacket momenta. We have verified that if we make initial wavepackets such that energy and momentum conservation cannot be simultaneously satisfied, no 4WM signal is observed. For instance, if we change the sign of the frequency difference between the two laser beams that comprise the second Bragg pulse, we will create a component with momentum $\mathbf{P}_3 = -2\hbar k\hat{x}$ instead of $\mathbf{P}_3 = 2\hbar k\hat{x}$. In this case there is no 4WM signal.

the three initial wavepackets separate, the nonlinear term will produce an additional wavepacket that satisfies the condition $\mathbf{P}_4 = \mathbf{P}_1 - \mathbf{P}_2 + \mathbf{P}_3 = \hbar k(\hat{x} - \hat{y})$, see Fig. 1a, as well as energy and particle-number conservation.

We have performed a two-dimensional numerical simulation of 4WM using equation (1) and the technique of ref. 9. The interaction energy (chemical potential) was chosen to be the same as it would be in three dimensions with a scattering length of 2.8 nm. The simulation releases 10^6 atoms from a trap with $\nu_x = 84$ Hz and $\nu_y = 59$ Hz. After 600 μ s, the condensate was projected into the three initial momentum states. Figure 2a shows the atomic density 1.8 ms after this projection. The most important feature of Fig. 2a is the new wavepacket of atoms with momentum $\mathbf{P}_4 = \mathbf{P}_1 - \mathbf{P}_2 + \mathbf{P}_3$ generated by 4WM. The 4WM peak does not appear when the nonlinear term is absent.

Figure 2b is a false-colour image showing the results of the experiment. The atoms were imaged 6.1 ms after the second Bragg pulse by optically pumping the atoms to the $3S_{1/2}$, $F = 2$ state, and absorption-imaging⁵ on the $3S_{1/2}$, $F = 2 \rightarrow 3P_{3/2}$, $F' = 3$ transition. The 4WM wavepacket is clearly visible. For this image, the numbers of atoms in each wavepacket were measured to be: $N_1 = 4.8(\pm 0.5) \times 10^5$, $N_2 = 5.3(\pm 0.5) \times 10^5$, $N_3 = 5.1(\pm 0.5) \times 10^5$ and $N_4 = 1.8(\pm 0.2) \times 10^5$, where the uncertainties are mainly due to uncertainties in background subtraction. The numbers of atoms in the three initial wavepackets N_1^0 , N_2^0 and N_3^0 can be deduced using particle number conservation: $N_1^0 = N_1 + N_4$, $N_2^0 = N_2 - N_4$, and $N_3^0 = N_3 + N_4$. Defining the 4WM efficiency to be $\epsilon = N_4/N$, where $N = \sum_{j=1}^3 N_j^0 = \sum_{j=1}^4 N_j$, we obtain a conversion efficiency of $10.6(\pm 0.13)\%$. This is the best we have observed, although under similar conditions we have also observed conversion efficiencies of only 6%. This difference suggests the influence of some uncontrolled experimental conditions, such as laser beam inhomogeneities, or non-zero average velocity of the released condensate. By comparison, the calculation of Fig. 2a gives an efficiency of 10%, albeit for only 10^6 atoms.

Equation (1) can be used to make a simple prediction about the expected nonlinear dependence of the 4WM signal on the numbers of atoms in the initial wavepackets. Substituting $\Psi = \sum_{j=1}^4 \Psi_j$ (where Ψ_j correspond to the individual momentum components) into equation (1), we find the initial rate of growth of the 4WM amplitude, $\partial \Psi_4 / \partial t \propto \Psi_1 \Psi_2^* \Psi_3$. We estimate the number of atoms in the fourth wave by multiplying this rate by a characteristic interaction time τ , proportional to the diameter of the condensate, squaring and integrating over space: $N_4 \propto n_1 n_2 n_3 V \tau^2$, where the density $n_j = N_j^0 / V$. In the Thomas–Fermi limit⁷, the volume of the condensate $V \propto N^{3/5}$, and $\tau \propto N^{1/5}$. Hence we expect $N_4/N \propto (N_1^0 N_2^0 N_3^0) N^{-9/5}$, a dependence which is supported by the numerical calculations. This nonlinear behaviour is clearly manifested in the initial linear growth seen in Fig. 3, where we vary the

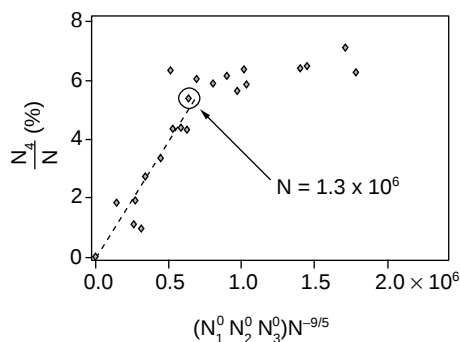


Figure 3 Measured conversion efficiency. Efficiency N_4/N is plotted as a function of $(N_1^0 N_2^0 N_3^0) N^{-9/5}$. The initial linear dependence is a signature of 4WM with matter waves. The dashed line is a fit to the first 12 points to guide the eye.

number of atoms in the original BEC and measure the number of atoms in the respective wavepackets. The data also show saturation at high N , as does the corresponding theory, although the maximum theoretical efficiency is somewhat higher.

We now reconsider Fig. 1b. Here the process is seen as degenerate 4WM (where the magnitudes of all momenta are equal) in a geometry equivalent to phase-conjugation in optics¹²: indeed $\mathbf{P}_4'$ is the momentum conjugate of $\mathbf{P}_2'$. So this can also be considered as a demonstration of phase conjugation with matter waves. As in the case of optical phase conjugation, the process would work regardless of the angle between $\mathbf{P}_1'$ and $\mathbf{P}_2'$ (90° in the present case). If one alters the first Bragg pulse by changing only the angle of $\mathbf{k}_2$ (and appropriately changing ν_2) the magnitude and direction of $\mathbf{P}_2$ in the laboratory frame are changed, so that in the moving frame only the angle of $\mathbf{P}_2'$ is changed.

We emphasize that just as optical 4WM requires coherent light sources to coherently build up the generated wave, a condensate is also crucial for coherent generation of matter waves. If atoms are above the Bose–Einstein condensation temperature, the number density is necessarily low and the phase-matching condition is different for each velocity class. Both dramatically diminish the 4WM conversion efficiency.

In spite of the strong analogy between atom and optical 4WM, there are fundamental differences. In optical 4WM, the energy–momentum dispersion relation is $E = [c n(k)] \hbar k$ (where $n(k)$ is the dispersive refractive index), whereas for massive particles (neglecting the matter-wave refractive index due to the atom–atom interaction energy) $E = P^2/2M$. Because atoms are neither created nor destroyed, the only 4WM processes allowed for matter waves conserve particle number. This is not the case for optical 4WM where, for example, in frequency tripling three photons are annihilated and one is created. Particle, energy and momentum conservation limit all matter 4WM processes to configurations that can be viewed as degenerate 4WM in an appropriate moving frame.

The present experiment used relatively large momenta. If we were to use momenta small enough to couple to phonons or other collective excitations of the condensate, we would be able to study these excitations and their nonlinear interactions with each other and with large-momentum excitations. We could also change the internal states by using Raman transitions⁴ to scatter atoms in one internal state from the matter-wave grating formed by atoms in a different internal state. It should even be possible to study 4WM between different isotopes or elements. Furthermore, just as nonlinear optics can create quantum correlations between photon beams, nonlinear atom optics may lead to the study of non-classical matter-wave fields. □

Received 29 December 1998; accepted 16 February 1999.

1. Franken, P. A., Hill, A. E., Peters, C. W. & Weinreich, G. Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961).
2. Maker, P. D. & Terhune, R. W. Study of optical effects due to an induced polarization third order in the electric field strength. *Phys. Rev.* **137A**, A801–A818 (1965).
3. Mewes, M.-O. *et al.* Output coupler for Bose–Einstein condensed atoms. *Phys. Rev. Lett.* **78**, 582–585 (1997).
4. Hagley, E. W. *et al.* A well collimated quasi-continuous atom laser. *Science* (in the press).
5. Anderson, M. H. *et al.* Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
6. Lens, G., Meystre, P. & Wright, E. W. Nonlinear atom optics. *Phys. Rev. Lett.* **71**, 3271–3274 (1993).
7. Dalfovo, F., Giorgini, S., Pitaevskii, L. P. & Stringari, S. Theory of trapped Bose–condensed gases. *Rev. Mod. Phys.* **71** (in the press).
8. Goldstein, E. V., Plattner, K. & Meystre, P. Atomic phase conjugation. *Quantum Semiclass. Opt.* **7**, 743–749 (1995).
9. Trippenbach, M., Band, Y. B. & Julienne, P. S. Four wave mixing in the scattering of Bose–Einstein condensates. *Opt. Express* **3**, 530–537 (1998).
10. Kozuma, M. *et al.* Coherent splitting of Bose–Einstein condensed atoms with optically induced Bragg diffraction. *Phys. Rev. Lett.* **82**, 871–875 (1999).
11. Petrich, W., Anderson, M. H., Ensher, J. R. & Cornell, E. A. Stable, tightly confining magnetic trap for evaporative cooling of neutral atoms. *Phys. Rev. Lett.* **74**, 3352–3355 (1995).
12. Shen, Y. R. *The Principles of Nonlinear Optics* 249–251 (Wiley, New York, 1984).

Acknowledgements. We thank K. Burnett, C. W. Clark, M. Kozuma and D. E. Pritchard for discussions. This work was supported in part by the US Office of Naval Research and NASA.

Correspondence and requests for materials should be addressed to L.D. (e-mail: lu.deng@nist.gov).

Vanishing of phase coherence in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$

J. Corson*, R. Mallozzi*, J. Orenstein*, J. N. Eckstein† & I. Bozovic‡

* Department of Physics, University of California, Berkeley, and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

† Department of Physics, University of Illinois, Urbana, Illinois 61801, USA

‡ Oxxel GmbH, Fahrenheit Str. 9, Technologiepark Universität, D-28359 Bremen, Germany

Although the binding of electrons into Cooper pairs is essential in forming the superconducting state, its remarkable properties—zero resistance and perfect diamagnetism—require phase coherence among the pairs as well. When coherence is lost at the transition temperature T_c , pairing remains, together with phase correlations which are finite in space and time. In conventional metals, Cooper pairs with short-range phase coherence survive no more than 1 K above T_c . In underdoped high- T_c copper oxides, spectroscopic evidence for some form of pairing is found up to a temperature T^* , which is roughly 100 K above T_c (refs 1–3). How this pairing and Cooper-pair formation are related is a central problem in high- T_c superconductivity. The nature of this relationship hinges on the extent to which phase correlations accompany pairing in the normal state⁴. Here we report measurements of high-frequency conductivity that track the phase-correlation time τ in the normal state of the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ family of underdoped copper oxide superconductors. Just above T_c , we find that τ reflects the motion of thermally generated topological defects in the phase, or vortices^{5,6}. However, vortex proliferation reduces τ to a value indistinguishable from the lifetime of normal-state electrons at 100 K, well below T^* .

In layered superconductors like the copper oxides, the density of paired electrons per layer, N_s , determines the phase stiffness, the energy cost to produce spatial variations in phase within a layer. Superconductivity persists until the thermal energy $k_B T$ (where k_B is

the Boltzmann constant) becomes comparable to the phase-stiffness energy $k_B T_\theta \equiv N_s \hbar^2 / m^*$, where m^* is the carrier effective mass. In conventional superconductors at low temperature, $k_B T_\theta$ is much greater than the pair binding energy. In such materials, the transition to the normal state is driven by Cooper-pair unbinding, which reduces N_s and T_θ becomes comparable to T .

In copper oxide superconductors the conventional ordering of the binding and phase stiffness energies is reversed⁴. As their carrier concentration is modified, away from the optimal value for highest T_c towards the Mott insulator (underdoping), this reversal becomes more profound; reduced phase stiffness is accompanied by larger binding energy energy^{2,7}. Thus, in underdoped copper oxides, electrons may remain tightly bound while long-range phase coherence vanishes at T_c (refs 4, 8). A state with paired electrons and short-range phase correlations was proposed⁴ to account for the anomalous properties of copper oxides above the transition temperature, between T_c and T^* .

This conjecture has led to the search for normal-state remnants of the infinite d.c. conductivity and perfect diamagnetism which exist below T_c . For the most part, these consequences of partial phase coherence have not been observed. It can be argued, however, that if the correlation times are sufficiently short, the expected enhancement above normal-state values may be unobservable by low-frequency probes.

The above argument motivates the use of high-frequency techniques to capture the short-timescale dynamics. A powerful probe of this type is the complex frequency dependent conductivity $\sigma(\omega)$. Below T_c , the conductivity measures the phase-stiffness energy directly:

$$\sigma(\omega) = i\sigma_Q(k_B T_\theta / \hbar \omega) \quad (1)$$

where $\sigma_Q \equiv e^2 / \hbar d$ is the quantum conductivity of a stack of planar conductors with interlayer spacing d . Equally important is the behaviour of $\sigma(\omega)$ expected above T_c in the presence of short-range phase correlations. At low frequency, σ will approach a real constant: the normal-state d.c. conductivity. At sufficiently high probing frequency, however, the phase-fluctuating state becomes indistinguishable from the superconducting state and σ is expected to approach equation (1). The crossover between these two limits occurs at a frequency $\Omega \equiv 1/\tau$, where τ is the phase-correlation time.

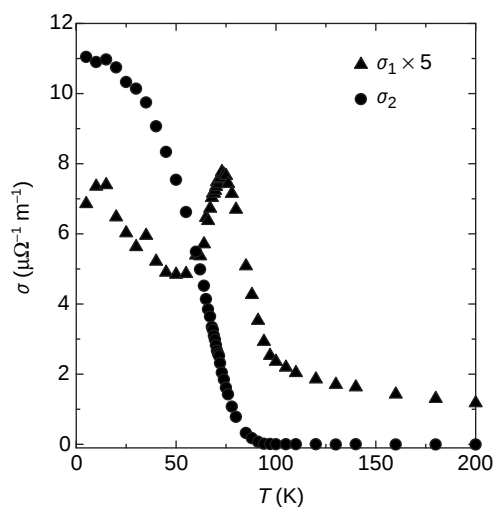


Figure 1 The complex conductivity σ measured at 100 GHz, as a function of the temperature T . The real part, σ_1 , is multiplied by five for ease of comparison with the imaginary part σ_2 . The real part is comprised of a peak near T_c superposed on a smooth background. We interpret the peak as the contribution from a partially coherent superfluid and the background as the contribution from quasiparticles.

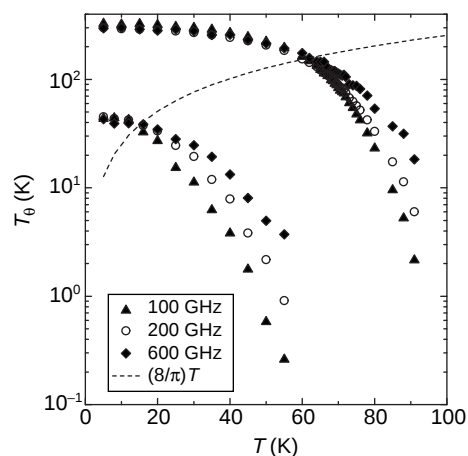


Figure 2 The dynamic (frequency dependent) phase-stiffness temperature, $T_\theta(\omega)$ as a function of temperature T . Data are shown for two samples, one with $T_c = 33$ K (left side) and the other with $T_c = 74$ K (right side). The three curves for each sample correspond to measurement frequencies of 100, 200 and 600 GHz. The family of curves identify a crossover from frequency-independent to frequency-dependent phase stiffness. The dashed line shows that the crossover corresponds to the KTB condition for two-dimensional melting, that is, when the phase stiffness and the temperature are related by $T_\theta = (8/\pi)T$.

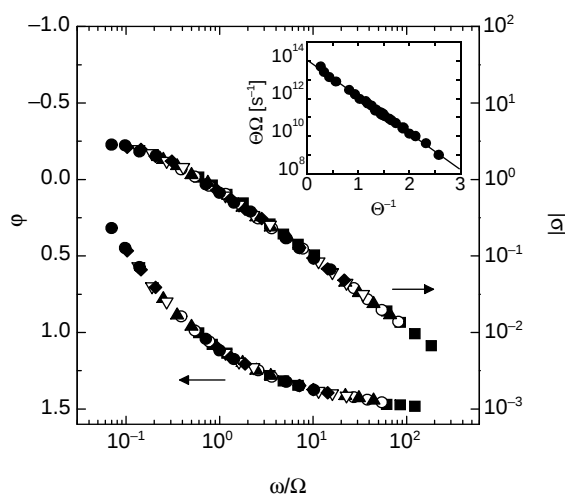


Figure 3 The conductivity phase angle, $\varphi = \tan^{-1}(\sigma_2/\sigma_1)$, and dimensionless conductivity magnitude $|\sigma|$, plotted as a function of the reduced frequency $\omega/\Omega(T)$. In calculating φ we include only the fluctuation contribution to σ_1 , which is obtained by subtracting the broad quasiparticle background from the peak seen in Fig. 1. The scales for φ , and $|\sigma|$, are shown on the left-hand, and right-hand axes, respectively. Each plot is comprised of data measured at temperatures in the range from 64 to 91 K, and frequencies from 100 to 400 GHz. Normalizing the frequency scale to the fluctuation frequency $\Omega(T)$ collapses the experimentally determined $\varphi(\omega, T)$ to a single curve. A similar data collapse as a function of $\omega/\Omega(T)$ is achieved when the conductivity magnitude is normalized by the quantity $\sigma_0 k_B T_0^0 / \hbar \Omega$. The values of $T_0^0(T)$ and $\Omega^{-1}(T)$ found by this scaling analysis are plotted in Fig. 4. The inset shows a remarkable exponential relationship between the fluctuation frequency Ω and the reduced phase-stiffness temperature $\Theta \equiv T/T_0^0$. The graphs shows $\Theta\Omega$ versus $1/\Theta$ on a semi-log scale. The line fit through the data corresponds to the exponential relation, $\Omega = (\Omega_0/\Theta)\exp(2C/\Theta)$, with $C = 2.23$ and $\Omega_0 = 1.14 \times 10^{14} \text{ s}^{-1}$.

Here we report $\sigma(\omega)$ measured on a set of underdoped copper oxide superconductors. We use the technique of time-domain transmission spectroscopy to capture the linear response at high frequency, from 100 to 600 GHz (ref. 9). Direct measurement of electric field rather than intensity yields both the real and imaginary parts of $\sigma(\omega)$ without Kramers–Kronig analysis. The samples are thin (400–655 Å) epitaxial films of underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (BSCCO) grown by atomic layer-by-layer molecular-beam epitaxy¹⁰. Underdoping is achieved by either varying oxygen concentration or replacement of calcium by dysprosium. Angle-resolved photoemission measurements as a function of carrier concentration have been performed on BSCCO samples grown in the same chamber⁷.

Figure 1 shows the real and imaginary parts of the conductivity, σ_1 and σ_2 respectively, at 100 GHz for an underdoped BSCCO film. In this sample, the resistive T_c was reduced to 74 K by replacement of 10% of calcium by dysprosium. The component related to dissipation, σ_1 , has a peak centred near T_c superposed on a rising background. We notice that σ_2 becomes observable near 100 K, more than 25 K above T_c , and has no obvious onset temperature. Both curves show that the transition from normal to superconducting state is broadened when viewed on a short timescale. Although partial phase coherence above T_c can give rise to such effects, more conventional interpretations must be considered as well. These include broadening due to static disorder, or gaussian fluctuations, in which the pairing amplitude fluctuates about zero and the phase is undefined.

To help distinguish these possibilities we examine the behaviour of σ_2 for different frequencies. For comparison with theory we convert σ_2 to the phase stiffness of an individual CuO_2 bilayer using equation (1), $k_B T_\theta(\omega) \equiv \hbar \omega (\sigma_2/\sigma_Q)$, where $\sigma_Q =$

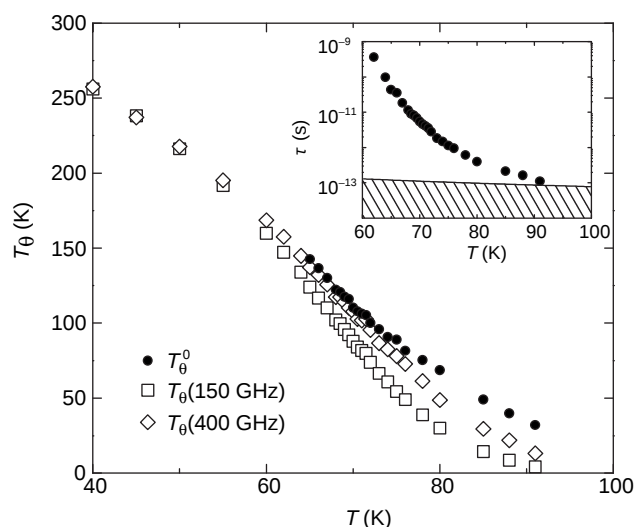


Figure 4 The phase correlation time $\tau = \Omega^{-1}$ and bare phase stiffness T_θ^0 found from the scaling analysis of conductivity data are plotted as a function of temperature. The main panel compares the bare stiffness, shown as filled circles, with the dynamic phase stiffness at 150 GHz (open squares) and at 400 GHz (open diamonds). The inset shows τ on a semi-log plot. The hatching defines a region where τ is less than the lifetime of carriers in the normal state, $\hbar/k_B T$. When the data points reach this region, phase fluctuations of the superconducting condensate become indistinguishable from the ballistic dynamics of normal electrons.

$1.58 \times 10^5 \Omega^{-1} \text{ m}^{-1}$ for a spacing between bilayers of 1.54 nm. Figure 2 shows $T_\theta(\omega)$ versus T on a semilog scale for the sample shown in Fig. 1 and for our most underdoped sample with $T_c = 33 \text{ K}$. This plot identifies a crossover in the dynamics with increasing temperature. At low T , the phase stiffness is frequency independent. With increasing temperature, T_θ begins to fan out for the different frequencies in our bandwidth, with the lowest-frequency data decreasing most rapidly. All five samples that we have measured show the behaviour described above. Furthermore, we find that for all samples the value of T_θ at the crossover, and the temperature at which it occurs, are related linearly. The dashed line in Fig. 2 shows how well the crossover is described by the simple relation $T_\theta = (8/\pi)T$.

The phase-stiffness dynamics shown in Fig. 2 suggest the relevance of the Kosterlitz–Thouless–Berezinskii (KTB) theory of two-dimensional melting^{5,6}. In the KTB picture, phase coherence is controlled by thermally generated vortices. In two dimensions, each vortex is a point defect around which the phase changes by 2π . In the ordered state vortices bind into pairs of opposite vorticity. The transition to the disordered state occurs when the first unbound vortices appear. Their random thermal motion leads to a finite phase-correlation time. Regardless of the microscopic details which distinguish one material from another, the transition is predicted to occur when the reduced temperature T/T_θ reaches a universal value of $\pi/8$.

The experimental signature of a KTB transition is the behaviour of the dynamical phase stiffness. At the critical temperature, T_{KTB} , the low-frequency T_θ drops discontinuously to zero from its value just below the transition, $(8/\pi)T_{\text{KTB}}$ (ref. 11). However, for $\omega > \Omega$, T_θ approaches the ‘bare’ phase stiffness of the underlying superconductivity, T_θ^0 , and varies smoothly through the transition. At

intermediate frequencies, $T_0(\omega)$ interpolates smoothly between these two limiting behaviours¹².

The behaviour seen in Fig. 2 is consistent with KTB dynamics, if we identify the crossover with T_{KTB} of an isolated bilayer. Above T_{KTB} , the conductivity is predicted to scale according to^{13–15}:

$$\frac{\sigma(\omega)}{\sigma_0} = \left(\frac{k_B T_0^0}{\hbar \Omega} \right) S(\omega/\Omega) \quad (2)$$

The scaling function $S(\omega/\Omega)$ is constrained by the physics of the high- and low-frequency limits. As $\omega/\Omega \rightarrow \infty$, S must approach $i\Omega/\omega$ in order for σ to assume its superconducting form, equation (1). At low frequencies, S approaches a real constant $S_1(0)$ which characterizes the d.c. conductivity of the normal state.

By comparing the measured complex conductivity to equation (2), we can extract both the phase stiffness and correlation time at each temperature. To analyse the experimental data in terms of equation (2), we note that the phase angle of the complex conductivity, $\varphi \equiv \tan^{-1}(\sigma_2/\sigma_1)$, equals the phase angle of $S(\omega/\Omega)$. Therefore φ depends only on the single parameter Ω , and is independent of T_0^0 . With the appropriate choice of $\Omega(T)$, all the measured values of φ should collapse to a single curve when plotted as a function of the normalized frequency ω/Ω . Knowing $\Omega(T)$, T_0^0 is obtained from a collapse of the normalized conductivity magnitude, $(\hbar\Omega/k_B T_0^0)|\sigma(\omega)|/\sigma_0$, to $|S(\omega/\Omega)|$. Figure 3 shows the collapse of the data to the phase angle and magnitude of S . As anticipated, S approaches a real constant in the limit $\omega/\Omega \rightarrow 0$, and approaches $i\Omega/\omega$ as $\omega/\Omega \rightarrow \infty$.

When analysed further, the data reveal a confirmation of thermal generation of vortices in the normal state. In the KTB picture we expect that the d.c. conductivity will equal $k_B T/n_f D \Phi_0^2$, which is the 'flux-flow' conductivity of n_f free vortices with quantized flux Φ_0 , and diffusivity D (ref. 16). Together with equation (2), this implies that Ω is a linear function of n_f that is, $\Omega = \Omega_0 n_f a_{vc}/\Theta$, where a_{vc} is the area of a vortex core, $\Theta \equiv T/T_0^0$ is the reduced temperature, and $\Omega_0 \equiv \pi^2 S_1(0) D/a_{vc}$. Moreover, we expect that n_f will be a thermally activated function, except for T very close to T_{KTB} . The activation energy is simply $C k_B T_0^0$, where C is a non-universal constant of order unity. It follows that the fluctuation frequency depends exponentially on the reciprocal of the reduced temperature, $\Omega = (\Omega_0/\Theta) \exp(-2C/\Theta)$.

The inset to Fig. 3 is a plot of $\log(\Theta\Omega)$ versus $1/\Theta$ which shows that the exponential relation is observed over nearly four orders of magnitude. This is direct evidence that vanishing of phase coherence in our samples reflects the dynamics of thermally generated vortices. From the slope and intercept of a straight-line fit we obtain $C = 2.23$ and $\Omega_0 = 1.14 \times 10^{14} \text{ s}^{-1}$.

In Fig. 4 we present the behaviour of the bare stiffness and phase-correlation time obtained from our measurement and modelling of $\sigma(\omega)$. The main panel contrasts T_0^0 with the dynamical stiffness $T_0(\omega)$ measured at 150 and 400 GHz. The inset shows τ as a function of temperature together with hatching that highlights the region where $\tau < \hbar/k_B T$.

The parameters displayed in Fig. 4 suggest that while phase correlations indeed persist above T_c , they vanish well below T^* . The loss of coherence is driven by the decrease of T_0^0 with increasing temperature, which renders the system increasingly defenceless to the proliferation of free vortices. This decrease of T_0^0 is consistent with a phenomenological description of a d -wave superconductor derived from a Mott insulator. In this picture, the phase stiffness of an underdoped copper oxide superconductor is undermined below T^* by the thermal generation of normal electrons very near the points in momentum space where the superconducting gap vanishes¹⁷. The pairing which remains in other regions of momentum space appears to contribute little to the overall phase stiffness. Near 95 K, τ falls to its minimum detectable value of $\hbar/k_B T$, which is the electron mean free time. Beyond this point, superconductivity becomes indistinguishable from the ballistic dynamics of normal

electrons and the recovery of the incoherent normal state is complete. □

Received 26 January; accepted 3 February 1999.

- Loesser, A. G. *et al.* Excitation gap in the normal state of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **273**, 325–329 (1996).
- Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high- T_c superconductors. *Nature* **382**, 51–54 (1996).
- Renner, Ch. *et al.* Pseudogap precursor of the superconducting gap in under- and over-doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. Lett.* **80**, 149–152 (1998).
- Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
- Berezinskii, V. L. Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group I. Classical systems. *Sov. Phys. JETP* **32**, 493–500 (1970); II. Quantum systems. *Sov. Phys. JETP* **34**, 610–617 (1972).
- Kosterlitz, J. M. & Thouless, D. J. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C* **6**, 1181–1203 (1973).
- Harris, J. M. *et al.* Anomalous superconducting state gap size versus T_c behavior in underdoped $\text{Bi}_2\text{Sr}_2\text{Ca}_{1-x}\text{Dy}_x\text{Cu}_2\text{O}_{8+\delta}$. *Phys. Rev. B* **54**, 15665–15668 (1996).
- Doniach, S. & Inui, M. Long-range Coulomb interactions and the onset of superconductivity in the high- T_c materials. *Phys. Rev. B* **41**, 6668–6678 (1990).
- Nuss, M. C. & Orenstein, J. in *Millimeter and Submillimeter Spectroscopy of Solids* (ed. Grüner, G.) 7–50 (Topics in Applied Physics Vol. 74, Springer, New York, 1998).
- Eckstein, J. N. & Bozovic, I. High-temperature superconducting multilayers and heterostructures grown by atomic layer-by-layer molecular beam epitaxy. *Annu. Rev. Mater. Sci.* **25**, 679–709 (1995).
- Kosterlitz, J. M. & Nelson, D. R. Universal jump in the superfluid density of two-dimensional superfluids. *Phys. Rev. Lett.* **39**, 1201–1205 (1977).
- Ambegaokar, V., Halperin, B. I., Nelson, D. R. & Siggia, E. D. Dynamics of superfluid films. *Phys. Rev. B* **20**, 1806–1826 (1980).
- Halperin, B. I. & Nelson, D. R. Resistive transition in superconducting films. *J. Low Temp. Phys.* **36**, 599–616 (1979).
- Minnhagen, P. The two-dimensional Coulomb gas, vortex unbinding, and superfluid-superconducting films. *Rev. Mod. Phys.* **59**, 1001–1066 (1987).
- Jonsson, A. & Minnhagen, P. Characteristics of two-dimensional vortex dynamics from XY-type models with Ginzburg-Landau dynamics. *Phys. Rev. B* **55**, 9035–9046 (1997).
- Bardeen, J. & Stephen, M. J. Theory of the motion of vortices in superconductors. *Phys. Rev. A* **140**, 1197–1207 (1965).
- Lee, P. A. & Wen, X. G. Unusual superconducting state of underdoped cuprates. *Phys. Rev. Lett.* **78**, 4111–4114 (1997).

Acknowledgements. This work was supported by the NSF, DOE and ONR. J.O. thanks the Aspen Center for Physics for the opportunity to discuss this work with participants of its high- T_c workshop.

Correspondence and requests for materials should be addressed to J.O. (e-mail: joeo@ux5.lbl.gov).

Aerosol-assisted self-assembly of mesostructured spherical nanoparticles

Yunfeng Lu*, Hongyou Fan*, Aaron Stump*, Timothy L. Ward*, Thomas Rieker* & C. Jeffrey Brinker*†

* University of New Mexico/NSF Center for Micro-Engineered Materials, The Advanced Materials Laboratory, 1001 University Boulevard SE, Albuquerque, New Mexico 87106, USA

† Sandia National Laboratories, Direct Fabrication Department, 1831, Albuquerque, New Mexico 87185, USA

Particles possessing nanometre-scale pores of well-defined size and connectivity are of interest for catalysis, chromatography and controlled release of drugs, and as fillers with low dielectric constant, pigments and hosts for optically active compounds^{1,2}. Silica containing ordered mesopores (of nanometre-scale width) can be prepared by templating of surfactant^{3,4} and block copolymer⁵ liquid-crystalline mesophases, and interfacial phenomena have been used to control the macroscopic form of these materials, providing mesoporous particles^{1,6}, fibres^{7,8} and films^{9,10}. A variety of spherical or nearly spherical particles has been reported^{1,6,7,11–13}, but the degree of ordering and the range of the porous mesostructures have been limited. Here we report a rapid, aerosol-based^{14–16} process for synthesizing solid, well-ordered spherical particles with stable pore mesostructures of hexagonal and cubic topology, as well as layered (vesicular) structures. Our method relies on evaporation-induced interfacial self-assembly¹⁷ confined to spherical aerosol droplets. This simple, generalizable process can be modified for the formation of ordered mesostructured thin films.

Schacht *et al.*⁶ first demonstrated that mesoscopically ordered hollow spheres could be prepared by interfacial reactions conducted in oil/water emulsions with varying extents of imposed shear. Huo *et al.*¹ extended this general emulsion-based approach to prepare spheres with diameters of 0.1–2.0 mm. Tanev and Pinnavaia¹² condensed silica in the interlayer regions of multilamellar vesicles to form roughly spherical particles with stable lamellar mesostructures. Bruinsma *et al.*⁷ used spray drying to prepare hollow spherical particles or collapsed, irregular particles. These various pathways to the preparation of nominally spherical mesostructured particles have so far resulted in irregular shapes and/or particles with non-uniform or ill-defined mesostructures.

Our process starts with a homogeneous solution of soluble silica plus surfactant prepared in an ethanol/water solvent with initial surfactant concentration c_0 much less than the critical micelle concentration c.m.c. Using the apparatus depicted in Fig. 1, we generate an aerosol dispersion within a tubular reactor. In a continuous, ~6-second process, the aerosol particles are dried, heated and collected. As recently demonstrated for thin-film formation¹⁷, preferential alcohol evaporation¹⁸ during drying enriches the particles in surfactant, water and silica, inducing micelle formation and successive co-assembly of silica–surfactant micellar species into liquid-crystalline mesophases. The resulting particles are commonly solid, with highly ordered hexagonal, cubic or vesicular mesostructures (Fig. 2). Through introduction of metal complexes or organic dyes/precursors into the starting solution, we can use the same process to prepare ordered nanocomposite particles (Fig. 3).

Figure 2 shows representative transmission electron microscope (TEM) images of calcined silica particles prepared using CTAB, Brij-58, Brij-56 or P123 templates (see Methods for details of templates). Figure 3 shows TEM images of various metal–, organic– and polymer–silica nanocomposite particles. CTAB produces particles exhibiting a highly ordered hexagonal mesophase over the concentration range $c_0 = 0.06$ – 0.16 M (Figs 2a and 3a). Many of the particles adopted a polyhedral shape that is hexagonal in cross-section, a macroscopic manifestation of the local packing geometry¹⁹. Use of the non-ionic surfactants (Brij-56/58) commonly resulted in vesicular mesostructures, but cubic and hexagonal mesostructures were also attainable (Figs 2b and 3b). Distinct from related bulk and thin-film lamellar structures that collapse on calcination²⁰, the three-dimensional connectivity of the

nested spherical shells comprising the vesicular mesophase mechanically stabilizes the structure against collapse during surfactant removal. Under processing conditions where the outer silica shells solidify before the drying shrinkage is complete, uniform dimpling of the shell surface results (Fig. 3c), in order to preserve the solidified shell surface area. This may be contrasted with the shrinkage-induced formation of toroidal or raisin-like collapsed morphologies commonly observed for spray-dried powders⁷.

Figure 4 shows small-angle X-ray scattering (SAXS) data for calcined mesostructured silica particles prepared with CTAB, Brij-56 and P123 surfactants, along with the corresponding N_2 sorption isotherms ('as-prepared' particles had no internal surface area accessible to N_2 owing to retention of the surfactant templates). The main peak is indexed as the [100]-reflection of the hexagonal (CTAB) and vesicular mesophases. SAXS and N_2 sorption show the expected increase in d -spacing and pore size with increasing surfactant volume/molecular mass (CTAB < Brij-56 < P123).

The robustness of our evaporation-induced interfacial self-assembly (EISA) process, combined with the short residence time of the particles in the furnace, enable us to prepare a variety of mesostructured organic/silica, metal/silica and even enzyme/silica composite particles while retaining the respective structure, function and bio-activity of the guest species (A.S. and R. Bhatia, unpublished results). Furthermore, through addition of organic monomers and thermal initiators to the parent solution, we can co-organize and polymerize mesostructured silica–polymer nanocomposite particles, as recently demonstrated for coatings²¹. The mesostructures shown in Fig. 3 illustrate the diversity of nanocomposite

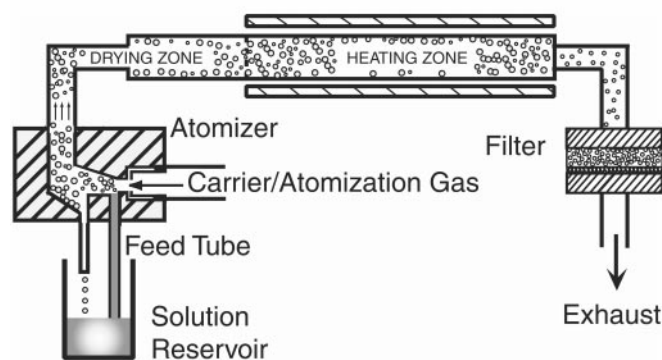


Figure 1 Diagram of aerosol reactor. Silica/surfactant aerosols were generated using a commercial atomizer (Model 3076, TSI, Inc., St Paul, MN) operated with nitrogen as a carrier/atomization gas. This atomizer produces aerosol droplets with a size distribution characterized by a geometric standard deviation of 2 (95% of the particles have diameters between 0.25 and 4 times the mean diameter). The pressure drop at the pinhole was 2.4 atm. Particles were collected on a Teflon filter. TEM grid, or silicon substrate maintained at 80 °C. For film experiments the furnace was not operated. For one experiment, HCl vapour was introduced into the reactor through a coaxial tube positioned at the centre of the heating zone.

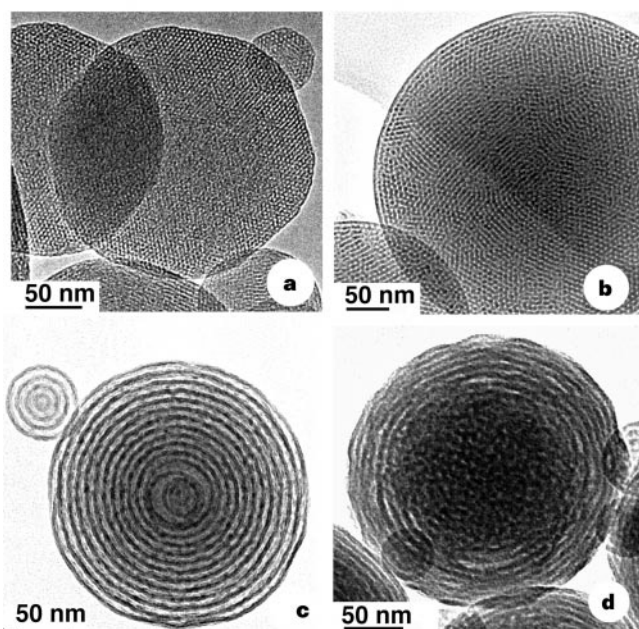


Figure 2 Representative TEM micrographs of mesostructured silica particles. **a**, Faceted, calcined particles with a hexagonal mesophase ($d_{100} = 32.5$ Å). The sol was prepared using 5 wt% CTAB as the surfactant template and molar ratio TEOS:EtOH:water:HCl:CTAB = 1:22:5:0.004:0.15. **b**, Calcined particles showing cubic mesostructure. The sol was prepared using 4.2 wt% Brij-58 as the surfactant template and molar ratio TEOS:EtOH:water:HCl:Brij-58 = 1:0:67:0.004:0.046. **c**, Calcined particles showing a vesicular mesophase ($d_{100} = 92$ Å). The sol was prepared using 5% P123 as the triblock copolymer template, and molar ratio TEOS:EtOH:water:HCl:P123 = 1:22:5:0.004:0.0096. **d**, Uncalcined silica particles showing 'growth' of ordered vesicular domains from the liquid–vapour interface. The particle interior has a disordered, worm-like mesostructure. The sol was prepared using 2.5% Brij-56 as the surfactant template, and molar ratio TEOS:EtOH:water:HCl:Brij-56 = 1:13:67:0.004:0.04. All calcination treatments were performed at 425 °C for 3 h in air (heating rate, 1 °C min^{−1}).

constructions and materials combinations attainable by our method.

Evaporation during aerosol processing creates within each droplet a radial gradient in surfactant concentration that steepens with time²² and maintains a maximum concentration at the droplet surface. Starting with an initially homogeneous solution (with c_0 less than the c.m.c.), the surfactant critical micelle concentration is exceeded first at the surface of the droplet, and, as evaporation proceeds, is progressively exceeded throughout the droplet. This surfactant enrichment induces silica-surfactant self-assembly into

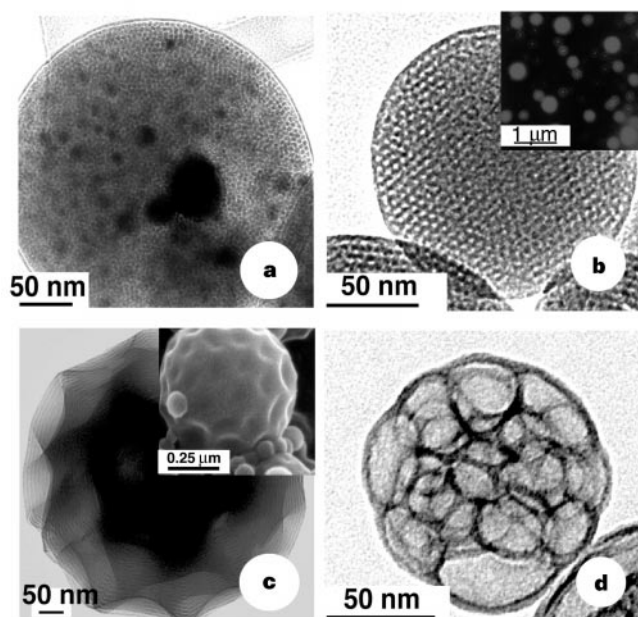


Figure 3 TEM micrographs of a variety of nanocomposite particles. **a**, Silica/gold nanocomposite particle with an ordered hexagonal mesostructure. Colloidal gold (particle size 1–3 nm) was prepared in inverse micelles according to the method of Brust *et al.*²⁸, using 1-dodecanethiol as a stabilizing agent. The gold colloids were added to a silica/5 wt% CTAB sol, prepared with molar ratio TEOS:EtOH:water:HCl:CTAB = 1:0.67:0.004:0.15. During aerosol processing, hydrophobic gold colloids partition into the hydrophobic micellar interiors and are subsequently incorporated in the hexagonal mesophase. TEM and X-ray diffraction indicate $d_{100} = \sim 60$ Å, (compared to 32.5 Å; Fig. 2a), consistent with incorporation of the gold colloids in the hexagonal silica mesophase. **b**, Fluorescent silica/rhodamine B nanocomposite particles exhibiting a hexagonal mesophase. The sol was prepared by adding 0.8 wt% rhodamine B to a silica/4.2 wt% Brij-56 sol prepared with molar ratio TEOS:EtOH:water:HCl:Brij-56 = 1:13:67:0.004:0.068. Inset, optical photomicrograph of the fluorescence emission imaged through a rhodamine B filter, confirming retention of the dye structure. Samples were exhaustively washed before optical microscopy. **c**, Silica/polymer nanocomposite particle showing pucker vesicular mesostructure (inset, SEM image showing uniform periodic dimpling). The sol was prepared by adding poly(propylene glycol dimethylacrylate) (PPO, molecular mass 660) and thermal initiator, 1,1'-azobis (1-cyclohexanecarbonitrile) (ACHN) to a silica/5 wt% P123 sol prepared with molar ratio TEOS:ethanol:water:HCl:P123:PPO:ACHN = 1:22:5:0.004:0.0096:0.025:0.009. During aerosol processing, the oligomer and initiator are incorporated into the hydrophobic portion of the template bilayers²¹ resulting in thermally initiated polymerization of PPO. **d**, Silica/polymer nanocomposite particle showing reticulated foam structure. The sol was prepared as in **c** but with molar ratio TEOS:ethanol:water:HCl:P123:PPO:ACHN = 1:13:67:0.004:0.0096:0.025:0.009. The greater water concentration compared to **c** resulted in micelle formation (in the starting sol) and incorporation of PPO and ACHN into the micellar interiors. Subsequent micelle coalescence combined with thermally initiated polymerization during aerosol processing resulted in the reticulated nanostructure.

micelles and further organization into liquid-crystalline mesophases. The radial concentration gradient and the presence of the liquid-vapour interface (which serves as a nucleating surface^{9,17,23}) causes ordered silica-surfactant liquid-crystalline domains to grow radially inward (Fig. 2d) rather than outwards from a seed²⁴. Crucial for the formation of solid, completely ordered particles is maintenance of a liquid, or liquid-crystalline, state throughout the course of the EISA process. Premature solidification would result in the formation of hollow particles⁷ and inhibit orderly self-assembly that must proceed with continual restructuring of the evolving silica-surfactant mesophase in order to accommodate drying shrinkage. Evidence for such a semi-solid state is the formation of continuous, uniform films on the collection surface when the furnace is maintained at low temperature (<150 °C). As shown in Fig. 5, film formation occurs via droplet coalescence. The resulting films are highly ordered, which is important for applications such as membranes, and can span holes hundreds of times greater in diameter than the film thickness.

We note that the curvature of the particle surface profoundly influences mesostructure development. This is evident in the comparison of CTAB-templated particles (Fig. 2a) with the corresponding mesostructured films formed on flat substrates by EISA during dip-coating¹⁷. Unlike films—which have flat liquid-vapour interfaces and show a progressive change in mesostructure (disordered $\rightarrow$ hexagonal $\rightarrow$ cubic $\rightarrow$ lamellar) with increasing surfactant concentration¹⁷—particles prepared with comparable CTAB

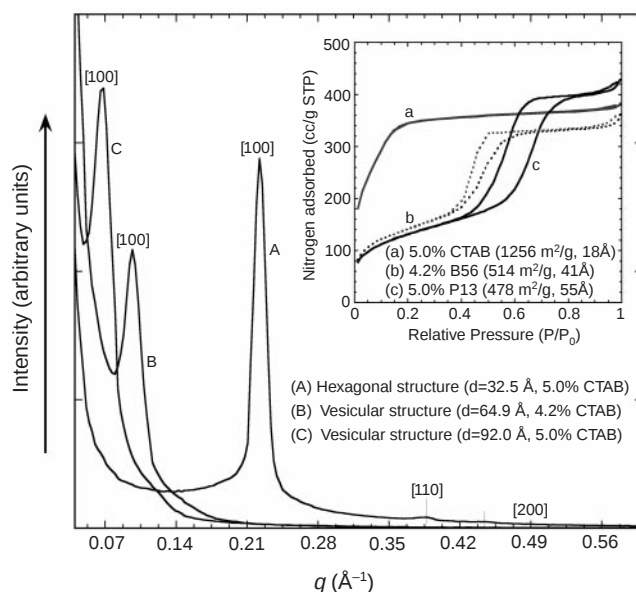


Figure 4 Small-angle X-ray scattering (SAXS) curves for silica particles with hexagonal (trace A) or vesicular (B and C) mesophases. Samples A and C were prepared as in Fig. 2a and c, respectively. Sample B was obtained for calcined particles showing a vesicular mesostructure ($d_{100} = 65$ Å). The sol was prepared using 4.2 wt% Brij-56 as the surfactant template, and molar ratio TEOS:EtOH:water:HCl: Brij-56 = 1:13:67:0.004:0.068. Inset, Corresponding N_2 adsorption-desorption isotherms, BET surface areas, and pore diameters. (We note that the calculated pore diameter of sample A is slightly less than 2.0 nm, the lower limit of mesoporosity according to the IUPAC definition. Thus the validity of the BET expression is questionable). SAXS experiments were performed at the University of New Mexico/Sandia National Laboratories SAXS facility²⁹. Samples (~ 0.5 mm thick) were prepared by sprinkling powder on adhesive cellophane tape used as a 1.5-cm-diameter window. SAXS data were collected (30 min per run) using the 5-m pinhole instrument in short geometry ($0.3 < q < 7$ nm⁻¹, where q is defined as $(4\pi/\lambda)\sin(2\theta/2)$, and 2θ is the scattering angle). SAXS data are uncorrected for background. Hydraulic pore diameters were calculated from the pore volume and BET surface area.

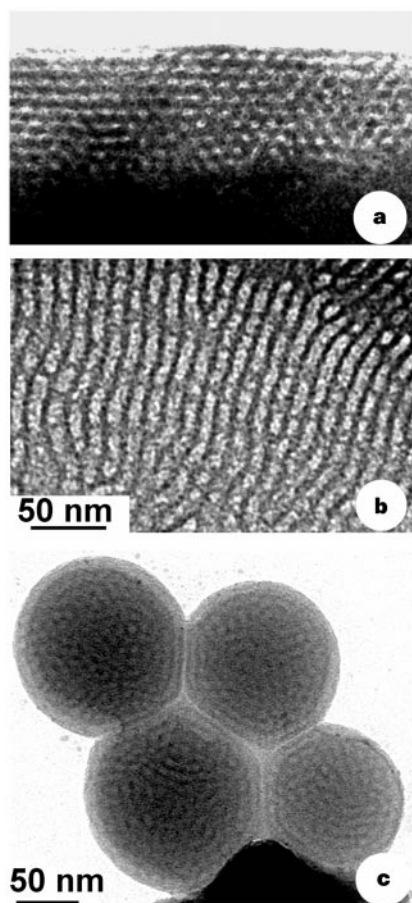


Figure 5 TEM images of mesostructured films formed by an aerosol-assisted thin film deposition route. A 5 wt% P123 sol (identical to that in Fig. 2c) was used. **a**, Cross-sectional TEM image of hexagonal mesostructured film deposited on silicon substrate. **b**, Plan-view image of hexagonal mesostructured film deposited on holey carbon grid. **c**, TEM image of corresponding aerosol droplets during coalescence on a holey carbon grid. For **a** and **b**, film samples were prepared using the apparatus shown in Fig. 1, but without operation of the furnace and with replacement of the filter with a silicon substrate (**a**) or holey carbon grid (**b**). The partially coalesced droplets (**c**) were prepared as in **b** but with introduction of HCl vapour to promote silica condensation and thereby 'freeze-in' the evolving mesostructure before complete coalescence. From **c** we conclude that film formation is possible by avoidance of particle solidification. The semi-solid nature of the particles (along with their low surface tensions) enables film formation over holes much greater in diameter than the film thickness.

concentrations show only disordered or hexagonal mesophases. We conclude that, because the liquid–vapour interface serves as a nucleating surface for liquid-crystal growth, the high curvature imposed by this interface alters the generally observed relationship between the surfactant packing parameter and the resulting mesostructure²⁵. Although CTAB commonly forms lamellar mesophases in bulk and thin-film samples^{3,20,26}, it appears that this molecule cannot pack into a cone truncated by surfaces of high and opposite curvature, as needed to direct the vesicular mesostructure. Only surfactants containing ethylene oxide (EO) blocks consistently gave vesicular mesophases. □

Methods

Precursor solutions were synthesized by the addition of cationic surfactant (CTAB; $\text{CH}_3(\text{CH}_2)_{15}\text{N}^+(\text{CH}_3)_3\text{Br}^-$), non-ionic surfactant (Brij-56; $\text{CH}_3(\text{CH}_2)_{15}-(\text{OCH}_2\text{CH}_2)_{10}-\text{OH}$; Brij-58; $\text{CH}_3(\text{CH}_2)_{15}-(\text{OCH}_2\text{CH}_2)_{20}-\text{OH}$) or triblock copolymers (Pluronic-P123; $(\text{EO})_{20}(\text{propylene oxide})_{70}(\text{EO})_{20}$) to an acidic silica sol (A2**). The acid concentration employed in the A2**

synthesis procedure was chosen to minimize the siloxane condensation rate²⁷, thereby promoting facile silica–surfactant self-assembly during aerosol processing. In a typical preparation, TEOS [$\text{Si}(\text{OCH}_2\text{CH}_3)_4$], ethanol, water and dilute HCl (mole ratios: 1:3.8:1:5 $\times 10^{-5}$) were refluxed at 60 °C for 90 min. The sol was diluted with ethanol (1:2) followed by addition of water and dilute HCl. Surfactants were added in the amounts needed to achieve initial surfactant concentrations c_0 ranging from 0.004 to 0.23 M. The final reactant mole ratios (TEOS:EtOH:H₂O:HCl:surfactant) were 1:(0–22):(5–67):0.004:(0.006–0.23). Sols used to prepare various nanocomposite particles were synthesized by addition of organosilates, organic dyes, polymers, metal colloids, or commercial catalysts directly to the precursor solutions.

Spherical mesostructured particles were prepared using an aerosol reactor (Fig. 1) operated at a volumetric flow rate of 2.6 l STP min⁻¹. Under these conditions, the flow is laminar (Reynolds number at 400 °C, 75) and the entrained aerosol particles experience ~3 s of drying at nominally room temperature followed by ~3 s of heating at 400 °C and finally collection on a filter maintained at 80 °C. The collected particles were characterized by TEM, SAXS, and N₂ sorption before and after calcination at 425 °C in air or N₂ (heating rate, 1 °C min⁻¹) to remove the surfactant templates.

Received 11 September 1998; accepted 4 January 1999.

- Huo, Q., Feng, J., Schuth, F. & Stucky, G. D. Preparation of hard mesoporous silica spheres. *Chem. Mater.* **9**, 14–17 (1997).
- Ozin, G. Nanochemistry: synthesis in diminishing dimensions. *Adv. Mater.* **4**, 612–649 (1992).
- Kresge, C., Leonowicz, M., Roth, W., Vartuli, C. & Beck, J. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Yang, H., Coombs, N. & Ozin, G. Mesoporous silica with micrometer-scale designs. *Adv. Mater.* **9**, 811–814 (1997).
- Zhao, D. *et al.* Triblock copolymer syntheses of mesoporous silica with periodic 50 and 300 Å pores. *Science* **279**, 548–552 (1998).
- Schacht, S., Huo, Q., Voigt-Martin, I. G., Stucky, G. D. & Schuth, F. Oil-water interface templating of mesoporous macroscale structures. *Science* **273**, 768–771 (1996).
- Bruinsma, P. J., Kim, A. Y., Liu, J. & Baskaran, S. Mesoporous silica synthesized by solvent evaporation: Spun fibers and spray-dried hollow spheres. *Chem. Mater.* **9**, 2507–2512 (1997).
- Yang, P., Zhao, D., Chmelka, B. F. & Stucky, G. D. Triblock-copolymer-directed syntheses of large-pore mesoporous silica fibers. *Chem. Mater.* **10**, 2033–2036 (1998).
- Yang, H., Coombs, N., Sokolov, I. & Ozin, G. Free-standing and oriented mesoporous silica films grown at the air–water interface. *Nature* **381**, 589–592 (1996).
- Aksay, I. *et al.* Biomimetic pathways for assembling inorganic thin films. *Science* **273**, 892–898 (1996).
- Grun, M., Lauer, I. & Unger, K. K. The synthesis of micrometer- and submicrometer-size spheres of ordered mesoporous oxide MCM-41. *Adv. Mater.* **9**, 254–257 (1997).
- Tanev, P. & Pinnavaia, T. Biomimetic templating of porous lamellar silicas by vesicular surfactant assemblies. *Science* **271**, 1267–1269 (1996).
- Yang, H., Vovk, G., Coombs, N., Sokolov, I. & Ozin, G. A. Synthesis of mesoporous silica spheres under quiescent aqueous acidic conditions. *J. Mater. Chem.* **8**, 743–750 (1998).
- Messing, G. L., Zhang, S.-C. & Jayanthi, G. V. Ceramic powder synthesis by spray pyrolysis. *J. Am. Ceram. Soc.* **76**, 2707–2726 (1993).
- Gurav, A., Kodas, T. T., Pluym, T. & Xiong, Y. Aerosol processing of materials. *Aerosol Sci. Technol.* **19**, 411–452 (1993).
- Pratsinis, S. E. & Mastrangelo, S. V. R. Material synthesis in aerosol reactors. *Chem. Eng. Prog.* **85**, 62–66 (1989).
- Lu, Y. *et al.* Continuous formation of supported cubic and hexagonal mesoporous films by sol-gel dip-coating. *Nature* **389**, 364–368 (1997).
- Jayanthi, G. V., Zhang, S. C. & Messing, G. L. Modeling of solid particle formation during solution aerosol thermolysis. *Aerosol Sci. Technol.* **19**, 478–490 (1993).
- Yang, H., Coombs, N., Dag, O., Sokolov, I. & Ozin, G. Free-standing mesoporous silica films; morphogenesis of channel and surface patterns. *J. Mater. Chem.* **7**, 1755–1761 (1997).
- Yang, H., Coombs, N. & Ozin, G. A. Morphogenesis of shapes and surface patterns in mesoporous silica. *Nature* **386**, 692–695 (1997).
- Israelachvili, J. *Intramolecular and Surface Forces* (Academic, San Diego, 1992).
- Bull, L. M. *et al.* In *Zeolites and Related Microporous Materials: State of the Art 1994* (eds Bull, L. M. *et al.*) 429–434 (Elsevier Science, Amsterdam, 1994).
- Brinker, C., Sehgal, R., Raman, N., Schunk, P. & Headley, T. Polymer approach to supported silica membranes. *Sol-Gel Sci. Technol.* **2**, 469–476 (1994).
- Brust, M., Walker, M., Bethell, D., Schiffrin, D. J. & Whyman, R. Synthesis of thiol-derivatized gold nanoparticles in a 2-phase liquid-liquid system. *J. Chem. Soc. Chem. Commun.* **7**, 801–802 (1994).
- Reiker, T. P. & Hubbard, P. F. The University of New Mexico/Sandia National Laboratories small-angle x-ray scattering laboratory. *Rev. Sci. Instrum.* (submitted).

Acknowledgements. We thank A. Singh and J. Nebo for discussions about porous particles, and A. Sellinger for discussions of oligomeric swelling agents. We also thank G. P. Lopez for assistance with the fluorescent emission microscopy and W. Gong for assistance with TEM. This work was partially supported by the UNM/NSF Center for Micro-Engineered Materials and the DOE Basic Energy Sciences Program. TEM investigations were performed in the Department of Earth and Planetary Sciences at the University of New Mexico. This work was done under contract from the US Department of Energy. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy.

Correspondence and requests for materials should be addressed to C.J.B. (e-mail: cjbrink@sandia.gov).

Molecular-sieve catalysts for the selective oxidation of linear alkanes by molecular oxygen

John Meurig Thomas, Robert Raja, Gopinathan Sankar & Robert G. Bell

Davy Faraday Research Laboratory, The Royal Institution of Great Britain, 21 Albemarle Street, London W1X 4BS, UK

Terminally oxidized hydrocarbons are of considerable interest as potential feedstocks for the chemical and pharmaceutical industry, but the selective oxidation of only the terminal methyl groups in alkanes remains a challenging task. It is accomplished with high efficiency and selectivity by some enzymes; but inorganic catalysts, although inferior in overall performance under benign conditions, offer significant advantages from a processing standpoint¹. Controlled partial oxidation is easier to achieve with 'sacrificial' oxidants, such as hydrogen peroxide², alkyl hydroperoxides or iodosylbenzene³, than with molecular oxygen or air. These sacrificial oxidants, themselves the product of oxidation reactions, have been used in catalytic systems involving tailored transition-metal complexes in either a homogeneous state^{4–6}, encapsulated in molecular sieves^{7–9} or anchored to the inner surfaces of porous siliceous supports¹⁰. Here we report the design and performance of two aluminophosphate molecular sieves containing isolated, four-coordinated Co(III) or Mn(III) ions that are substituted into the framework and act, in concert with the surrounding framework structure, as regioselective catalysts for the oxidation of linear alkanes by molecular oxygen. The catalysts operate at

temperatures between 373 K and 403 K through a classical free-radical chain-oxidation mechanism. They are thus able to use molecular oxygen as oxidant, which, in combination with their good overall performance, raises the prospect of using this type of selective inorganic catalyst for industrial oxidation processes.

Cobalt(II) is one of the transition-metal ions that, when occupying a small percentage of the framework (tetrahedral) sites in a molecular sieve, may be raised^{11–13} to a higher oxidation state (Co(III)) while remaining within the framework as active sites for the catalytic oxidation in air of cyclohexane¹⁴ and other alkanes¹⁵. By taking advantage of this fact and choosing a molecular sieve that allows only end-on entry of the linear alkanes into the cavities containing the active sites (compare ref. 16) we have been able to design effective catalysts that favour functionalization by oxygen at the terminal CH₃ and penultimate CH₂ groups and operate with air as the oxidant. (We note that in conventional autoxidation of *n*-alkanes under homogeneous conditions and in the absence of steric constraints the regioselectivity is governed by the relative bond-dissociation energies. That is, whereas the bond dissociation energies decrease from 104 kcal mol^{–1} to 94.6 kcal mol^{–1} to 91 kcal mol^{–1} in going from primary to secondary to tertiary carbon atoms¹⁷, the corresponding selectivities increase.)

We selected the molecular sieve known as aluminophosphate (AIPO) number 18 (ref. 18; idealized formula Al₂₄P₂₄O₉₆), which has pores similar to the zeotypic analogue of the aluminosilicate mineral chabazite. A few atom per cent of various divalent metal ions may be readily accommodated into this material's framework tetrahedral sites, as isomorphous replacements for Al(III) ions. This is done by templated hydrothermal syntheses¹⁸. The template is driven off completely by heating in oxygen to ~550 °C, during which process the redox ions (Co(II) or Mn(II)) are raised to their 3+ oxidation states. In the course of the partial oxidations (see Methods) the transition-metal ions, which replace ~4 atom per

Table 1 Oxidation of *n*-alkanes in air: comparison of catalysts and product selectivity

Alkane	Catalyst	Conv. mole %	TON§	Product distribution (mol %)										Prim. sel. index‡	Ref.
				C ₁ -ol*	C ₁ -ald	acid	C ₂ -ol	C ₂ -one	C ₃ -ol	C ₃ -one	C ₄ -ol	C ₄ -one	other†		
<i>n</i> -C ₆ H ₁₂	CoAIPO-18	8.8	174.6	32.5	–	–	–	43.7	–	21.1	–	–	2.7	0.48	p.w.¶
<i>n</i> -C ₆ H ₁₂	CoAIPO-36	5.5	108.9	5.3	–	–	–	33.3	–	59.9	–	–	1.5	0.06	p.w.
<i>n</i> -C ₆ H ₁₂	MnAIPO-18	8.2	151.4	38.6	–	–	–	48.3	–	11.8	–	–	1.3	0.63	p.w.
<i>n</i> -C ₆ H ₁₂	MnAIPO-36	5.9	115.7	–	–	–	–	21.5	–	76.7	–	–	1.8	–	p.w.
<i>n</i> -C ₆ H ₁₄	CoAIPO-5	2.4	41.04	1.3	4.1	3.2	6.0	10.1	41.0	30.4	–	–	3.9	0.13	p.w.
<i>n</i> -C ₆ H ₁₄	CoAIPO-11	3.6	61.56	7.4	7.8	4.2	17.3	17.1	13.9	29.9	–	–	2.4	0.32	p.w.
<i>n</i> -C ₆ H ₁₄	CoAIPO-18	7.2	121.4	4.5	3.3	53.5	13.9	22.2	–	–	–	–	2.6	2.1	p.w.
<i>n</i> -C ₆ H ₁₄	CoAIPO-36	5.2	88.92	8.7	3.8	10.2	9.2	19.8	14.5	30.6	–	–	3.2	0.39	p.w.
<i>n</i> -C ₆ H ₁₄	MnAIPO-18	8.7	149.1	9.9	12.5	43.1	8.4	23.3	–	–	–	–	2.8	2.5	p.w.
<i>n</i> -C ₆ H ₁₄	MnAIPO-36	5.4	81.5	–	–	–	7.9	19.3	31.5	39.7	–	–	1.6	0	p.w.
<i>n</i> -C ₈ H ₁₈	CoAIPO-18	6.2	79.65	14.4	18.8	27.1	12.6	18.2	–	6.4	–	–	2.5	3.0	p.w.
<i>n</i> -C ₈ H ₁₈	CoAIPO-36	5.8	74.75	4.9	7.5	–	5.5	9.9	8.4	16.5	17.9	27.0	2.4	0.28	p.w.
<i>n</i> -C ₈ H ₁₈	MnAIPO-18	6.8	82.8	9.4	22.1	30.5	12.8	23.5	–	–	–	–	1.7	3.2	p.w.
<i>n</i> -C ₈ H ₁₈	MnAIPO-36	5.9	72.25	2.5	3.1	1.7	3.5	5.0	12.5	19.9	18.4	32.5	0.9	0.15	p.w.
<i>n</i> -C ₆ H ₁₂	MnTTTPPP (OAc)‡‡				10		75		15		–	–	–	0.11	4¶
<i>n</i> -C ₆ H ₁₄	MnTTTPPP (OAc)				19		62		18		–	–	–	0.31	4¶
<i>n</i> -C ₆ H ₁₄	U-cy-P-450				10.8		75.7		13.5		–	–	–	0.16	27#
<i>n</i> -C ₆ H ₁₄	I-cy-P-450				4.5		49.1		46.4		–	–	–	0.06	27☆
<i>n</i> -C ₇ H ₁₆	MnTTTPPP (OAc)	130**			26		52		17		5	–	–	0.59	4¶
<i>n</i> -C ₈ H ₁₈	MnTTTPPP (OAc)				21		48		16		15	–	–	0.53	4¶
<i>n</i> -C ₈ H ₁₈	Fe/Pd-zeolite A				21		33		26		20	–	–	0.54	5††
<i>n</i> -C ₈ H ₁₈	ω-hydroxylase	40 × 10 ³ μmol			90		10		–		–	–	–	–	28

Reaction conditions: substrate, ~50 g; catalyst, 0.5 g; oxidant (air), 1.5 MPa; temperature 373 K; time, 24 h.

* C₁-ol, 1-alcohol; C₁-al, 1-aldehyde; C₂-ol, 2-alcohol; C₂-one, 2-ketone; C₃-ol, 3-alcohol; C₃-one, 3-ketone; C₄-ol, 4-alcohol; C₄-one, 4-ketone.

† Mainly CO₂, CO, water and lower olefins/hydrocarbons in the gas phase.

‡ Primary selectivity index defined as the ratio of concentration of primary products to secondary/tertiary products normalized for the respective number of hydrogens in the alkane molecule.

§ Moles of substrate converted per mole of cobalt or manganese in the catalyst.

¶ Present work.

‡ C₆H₅IO used as the oxidant (exact conversion values were not available in the present format).

Rat liver microsomal cytochrome P-450 (uninduced).

☆ On treatment with phenobarbital different and less selective isozymes of cytochrome P-450 are induced.

** Yields based on metalloporphyrin (yields based on oxidant were not given as excess oxidant was used).

†† Gaseous mixture of H₂O₂ was used as the oxidant (*in situ* generation of hydrogen peroxide).

‡‡ Manganese 5, 10, 15, 20-tetrakis (2', 4', 6' triphenyl phenyl) porphyrin.

cent of the Al(III) ions, return to their 2+ states, and the catalyst changes colour from green $\leftrightarrow$ blue (for Co(III) $\leftrightarrow$ Co(II)) and blue $\leftrightarrow$ white (for Mn(III) $\leftrightarrow$ Mn(II)). Combined *in situ* extended X-ray absorption spectroscopy and X-ray powder diffraction^{11,19} show that the oxidation of the ion is accompanied by only minor structural changes, and these measurements yield quantitative structural information about the environment of the Co(III)- and Mn(III)-centred active sites (see Supplementary Information).

Measurements of the shifts in the K-edges and of the extended X-ray absorption fine structure of the Co and Mn ions show that these ions in the as-prepared (templated) form of CoAlPO-18 and MnAlPO-18 may be completely converted in dry air or O₂ to the

Co(III) or Mn(III) states, respectively. For reasons that are not yet clearly understood, only a fraction of the Co(II) or Mn(II) ions isomorphously incorporated into the AlPO-36, AlPO-11 and AlPO-5 structures are convertible to the (III) oxidation state by similar treatment in O₂ or dry air¹². (In MeAlPO-36 the fraction is $\sim$ 0.45, in MeAlPO-11, $\sim$ 0.30 and in MeAlPO-5, $\sim$ 0.20, where Me is Co or Mn, all these values being deduced¹² from changes in Co–O and Mn–O distances.)

The catalytic performance of the CoAlPO-18, MnAlPO-18, MgAlPO-18, CoAlPO-36, MnAlPO-36, MgAlPO-36, CoAlPO-11 and CoAlPO-5 molecular sieves was tested for the selective oxidation of *n*-pentane, *n*-hexane and *n*-octane (Table 1 and Fig. 1). As

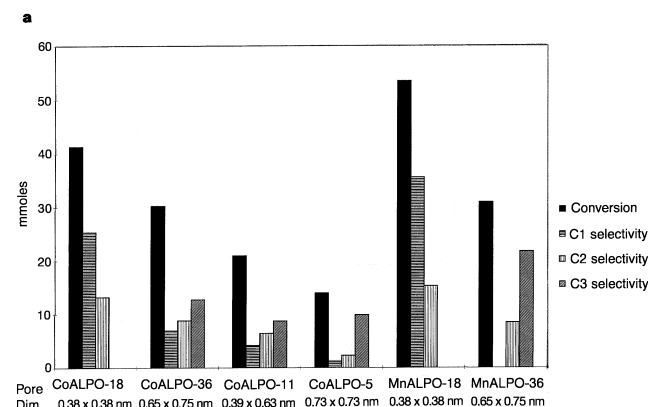
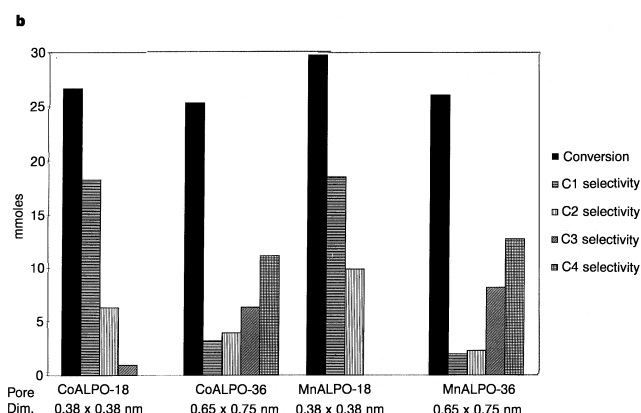


Figure 1 The performance of Co and Mn substituted molecular-sieve catalysts. The activity and various degrees of regioselectivity of the CoAlPO-18, CoAlPO-36, CoAlPO-11, CoAlPO-5, MnAlPO-18 and MnAlPO-36 in the oxidation of *n*-hexane (a) and *n*-octane (b) after 24 h, at 373 K and 1.5 MPa of dry air, are summarized. Whereas only the C₁ and C₂ positions were preferentially oxidized using the AlPO-



18 catalysts, the relatively larger pore dimensions of the AlPO-5 and AlPO-36 favoured oxidation at the C₃, C₄ and C₆ positions. After the reaction, both the CoAlPO-18 and MnAlPO-18 catalysts were washed thoroughly with methanol and activated at 550 °C for 16 h. They were then re-used at least twice without any significant decrease in activity or selectivity.

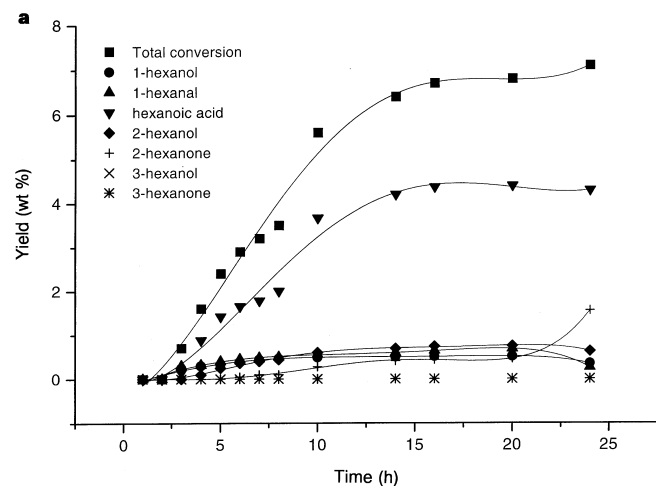
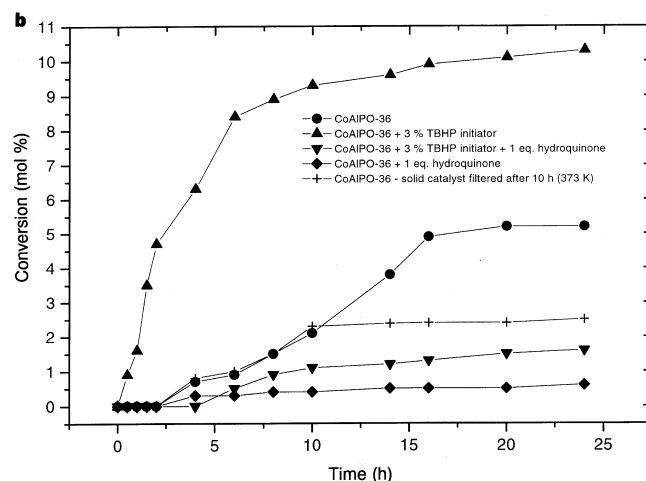


Figure 2 Kinetics of *n*-hexane oxidation in air over Co substituted AlPO catalysts. a, Typical kinetic plots for the oxidation of *n*-hexane over CoAlPO-18 catalyst under the conditions given in Methods. The kinetics of oxidation of *n*-hexane shows that the conversion increases linearly with time and levels off after $\sim$ 15 h. At this stage, $\sim$ 90% of the oxygen in the reactor is consumed, which could be the reason for the inhibition of the catalytic reaction. It is also possible that the alcohols and acids formed could also inhibit the reaction, owing to the hydrophilic nature of aluminophosphate molecular sieves. b, Kinetics of oxidation of *n*-hexane in air using CoAlPO-36 in the presence (upright triangles) and absence (circles) of initiator (3% *tert*-butyl hydroperoxide, TBHP). The addition of free-radical initiator greatly reduces the initial induction period, accelerates the rate of the catalytic reaction and gives rise to enhanced activity. (There was no change in



the product selectivity pattern when compared with the system without initiator.) Addition of small amounts of hydroquinone (a free-radical scavenger) completely suppresses the reaction (inverted triangles, diamonds), suggesting the participation of free radicals in the catalytic process. In an identical experiment, the solid catalyst was filtered off from the reaction mixture (when hot) after 10 h, and the reaction was continued for a further period of 14 h (plus signs). No increase in conversion was observed, indicating that the leached-out metal ions (if any) are not responsible for the observed activity. This experiment also rules out the possibility of any gas-phase radical reactions. The resulting filtrate was independently analysed by inductively coupled plasma (ICP) (and atomic absorption spectroscopy) for free or dissolved metal ions, and only trace amounts (<3 p.p.b.) were detected.

indicated by the primary selectivity index (Table 1), the overall performance of CoAlPO-18 and MnAlPO-18 is superior to that of all previously reported inorganic catalysts for the regioselective oxidation of linear alkanes, even though the latter use sacrificial oxidants, not air or O₂. An exception is the catalyst developed by Tolman *et al.*¹⁶, which uses a gas mixture of H₂ and O₂ to generate H₂O₂ *in situ* over finely divided Pd/Fe bimetallic particles.

Given the pore dimensions and the fact that, under reaction conditions, the framework transition-metal ions are all in their (III) state, it is unsurprising that CoAlPO-18 (Fig. 2a) and MnAlPO-18 show the greatest activity, being twice as active (in *n*-hexane oxidation) as a CoAlPO-11 catalyst²⁰ (which is effective in the low conversion autoxidation of cyclohexane). Our catalysts also show a high degree of regioselectivity: ~65.5% of the products of oxidation of *n*-hexane after 24 h are the result of oxyfunctionalization at the terminal methyl group for MnAlPO-18 (61.3% for CoAlPO-18). As much as 31.7% (36.1% for CoAlPO-18) consist of 2-hexanol and 2-hexanone, with no sign of attack of the C₃ carbon in the hexane or at the C₄ of *n*-octane (Fig. 1).

The results from the oxidation of *n*-pentane are different from those of *n*-hexane or *n*-octane. After 24 h, the C₃ and C₂ ketones are the main products from *n*-pentane oxidation (while C₁ hydroxylated products dominate in the case of *n*-hexane or *n*-octane). But at lower contact times (8 h; data not given in the table) we do observe 2- and 3-pentanols in the reaction mixture, which are rapidly converted to their corresponding ketones; small amounts of 1-pentanol are observed only after prolonged contact time. Equimolar mixtures of *n*-hexane and cyclohexane were also subjected to catalytic oxidation over either CoAlPO-18 or MnAlPO-18. In each case there was no conversion of the cyclohexane—it is too large to diffuse through the apertures of these catalysts. (Using either CoAlPO-36 or MnAlPO-36 as catalysts, cyclohexane was oxidized in air with relatively high turnover numbers²¹ (170–190 after 24 h) selectively to cyclohexanol, cyclohexanone and adipic acid.)

It is well known²² that the oxidation of alkanes by transition

metals generally involves the participation of free radicals; we believe, for three reasons, that the overall selective oxidations of all the alkanes reported here and that of cyclohexane reported elsewhere²¹ are no exceptions. Because of the small aperture of the AlPO-18 structure, it is easier to test the role of free radicals using the AlPO-36 structure which has apertures large enough to allow alkylhydroperoxy radicals and radical scavengers to enter their interior surfaces. First, there is an induction period in the yield versus time plots (typified by that for *n*-hexane, using CoAlPO-36 as the catalyst, Fig. 2b), and this period is greatly diminished by the prior addition of traces of free-radical initiators such as *tert*-butyl hydroperoxide (TBHP). Second, addition of small quantities of hydroquinone, a free-radical scavenger, greatly hinders the oxidation (see curves plotted with diamonds and inverted triangles, Fig. 2b). Third, in the oxidation of cyclohexane (the peroxide of which is easier to analyse, than that of *n*-alkanes) under identical conditions over a range of Co(III)- and Mn(III)-AlPO catalysts, the presence of cyclohexyl peroxide (which subsequently decomposes to cyclohexanol and cyclohexanone) has been directly detected²¹.

Although the precise mechanism of the oxidation of the alkanes involving CoAlPO and MnAlPO catalysts and dioxygen is not clear—as in the case of cytochrome P450²³ and other enzymatic reactions—there is little doubt that catalytic autoxidation^{1,16} dominates here. It is certain that framework-isolated Co(II) or Mn(II) ions present initially in the as-prepared sieve, when converted to the Co(III) or Mn(III) state, are vital for catalysis. It is likely that scope for expansion of the coordination shell of the transition-metal ion (as with the Mn(III) ion in the superoxide dismutase²⁴) is a significant factor in facilitating reaction. Divalent ions that cannot be raised to higher states of oxidation, such as Mg(II), are totally inactive as oxidation catalysts. Experiments using MgAlPO-18 and MgAlPO-36 as catalysts (for the oxidation of *n*-hexane under the conditions of Fig. 2 at 373 K) showed no activity. This fact also rules out the possibility that free radicals produced solely in the gas phase are significant in the autoxidation. Moreover, introduction of acetonitrile (known²⁵ to co-ordinate Co(II) ions) before the admission of air poisons the CoAlPO-18 catalyst.

Overall conversions in the oxidations reported here have to be kept to a low (~7%) level to avoid leaching of the transition metals by polar products (that is, carboxylic acids). However, they are comparable⁸ to those in industrially significant oxidations of cyclohexane, and superior to what has been achieved to date using other inorganic catalysts (see Table 1). The rates of reaction are rather slow, reflecting the diffusional requirements that have to be overcome when liquid reactants and products compete with one another and with gaseous oxygen for access to the sub-surface active sites. It is possible that further enhancement of intrinsic catalytic activity may be achieved by optimizing such variables as the concentration of Co(III) or Mn(III) ions substituted into the chabazitic cages, the crystallite size and the surface area. Moreover, by using a high-pressure flow system, optimized with respect to contact time and oxygen pressure, a higher yield of desirable products would be facilitated. It is also likely that other transition-metal ions may serve equally well or better as selective oxidation catalysts. But the regioselectivity of the oxidation is already remarkable, and may be understood from a computer simulation which determines²⁶ the state of lowest energy adopted by the alkane inside the chabazitic cavity of the AlPO-18 framework (Fig. 3). A bound hexane (or any other linear alkane) molecule, with its slightly bent end, gains ready access to the active sites, which selectively favour oxidation at the terminal carbons. □

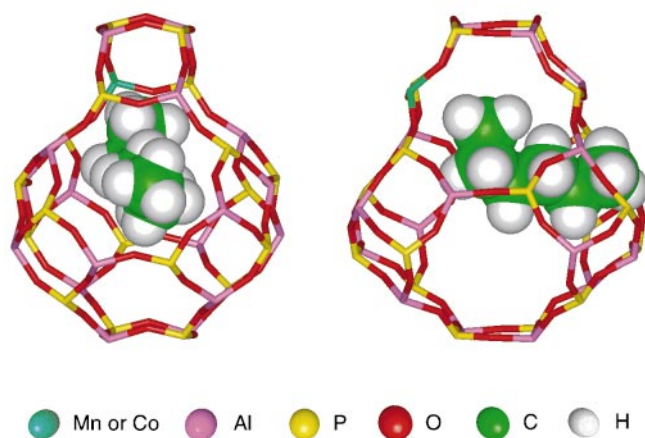


Figure 3 Energy-minimized configuration (two views) adopted by *n*-hexane at 0 K inside an AlPO-18 framework. This configuration was derived from a calculation that combines Monte Carlo, molecular dynamics, and docking procedures as described in ref. 26; the framework composition was assumed to be SiO₂, for which more reliable potentials are available than for AlPO₄. We note that the terminal methyl group (C₁) is significantly closer to a tetrahedral framework site than either C₂ or C₃, the respective distances being C₁ = 3.53 Å, C₂ = 3.75 Å and C₃ = 4.03 Å. This implies that Co(III) in the framework of the AlPO-18 cage, along with intracage O₂, exert a greater influence on C₁ than on other carbons in the hexane backbone.

Methods

Precursors. Catalyst precursors (Co(II) and Mn(II) framework-substituted AlPOs) were prepared as previously described^{12,18}.

Catalytic reactions. The oxidation reactions were performed in a high-pressure stainless-steel catalytic reactor (Cambridge Reactor Design) lined with

poly ether ether ketone (PEEK). Typically, 0.5 g of the catalyst—calcined to remove the occluded organic template and to convert Mn(II) and Co(II) ions to their (III) states using dry air at ~550 °C—was added to ~50 g of the substrate (*n*-pentane, *n*-hexane or *n*-octane), 4.8 g of 1,2 dichlorobenzene (internal standard), and stirred at a constant speed of 400 r.p.m. Dry air was fed under pressure into the reaction vessel (1.5 MPa), and the reactor was then sealed and heated to the desired temperature (373 K). Small aliquots of the sample were removed at regular intervals to study the kinetics of the reaction in detail—this was achieved using a liquid sampling valve and without perturbing the pressure in the reactor.

Product analysis. The products of the oxidations were analysed by gas chromatography (GC, Varian, Model 3400 CX) employing a BPX5 capillary column (25 m × 0.32 mm) and flame ionization detector. The total conversion and product distribution (estimated in moles) were evaluated with a calibration curve (obtained by injecting known amounts of authenticated standard compounds), the individual yields being calculated and normalized with respect to the GC response factors. The acids formed were esterified using BF₃ + CH₃OH and analysed as methyl esters⁸. The identity of the products was confirmed by injecting authenticated standard samples and further by GC/MS (Hewlett Packard). The mass balance was estimated at the end of the experiment, and the estimated error in analysis arising due to sampling and handling losses was between 3 and 5 mol%.

X-ray absorption studies. Mn and Co K-edge X-ray absorption spectra (XAS) were obtained *in situ* at Station 8.1 of the Daresbury Synchrotron radiation source (which operates at 2 GeV with a typical current of 150–250 mA). This station was equipped with Si(220) double crystal monochromator, ion chambers for measuring the incident-beam and transmitted-beam intensity, and a 13-element fluorescence detector (Canberra). The room-temperature data were collected after calcining the as-prepared material in pure oxygen at 550 °C employing an *in situ* cell described elsewhere¹². In addition, time-resolved *in situ* combined XRD/XAS data were recorded¹¹ at Station 9.3, which is also equipped with ion chambers and a 13-element fluorescence detector. XRD data were recorded below the Mn K-edge at a wavelength of 1.9375 Å for 180 s during the calcination of the catalyst in oxygen. XAS data were recorded for 380 s. The total time taken for the sequential XRD and XAS measurements is 10 min, which includes the dead time of 40 s associated with the movement of the monochromator. XAS data were analysed using the EXCALIB, EXBROOK and EXCURV 98 (ref. 29) suite of programs.

Received 3 July 1998; accepted 5 January 1999.

- Hill, C. L. (ed.) *Activation and Functionalisation of Alkanes* Ch. 6–8 (Wiley, Chichester, 1989).
- Tatsumi, T., Nakamura, M., Negishi, S. & Tominaga, H. Shape-selective oxidation of alkanes with H₂O₂ catalysed by titanosilicate. *J. Chem. Soc. Chem. Commun.* 476–477 (1990).
- Smegal, J. A. & Hill, C. L. Hydrocarbon functionalisation by the (iodosylbenzene) manganese(IV) porphyrin complexes from the (tetraphenylporphyrinato) manganese(III)-iodosylbenzene catalytic hydrocarbon oxidation system. Mechanism and reaction chemistry. *J. Am. Chem. Soc.* **105**, 3515–3521 (1983).
- Cook, B. R., Reinert, T. J. & Suslick, T. S. Shape selective alkane hydroxylation by metalloporphyrin catalysts. *J. Am. Chem. Soc.* **108**, 7281–7286 (1986).
- Mansuy, D. Cytochrome-P450 and synthetic models. *Pure Appl. Chem.* **59**, 759–770 (1987).
- Lyons, J. E., Ellis, P. E. Jr & Myers, W. K. Jr Halogenated metalloporphyrin complexes as catalysts for selective reactions of acyclic alkanes with molecular oxygen. *J. Catal.* **155**, 59–73 (1995).
- Herron, N. & Tolman, C. A. A highly selective zeolite catalyst for hydrocarbon oxidation—a completely inorganic mimic of the alkane omega-hydroxylases. *J. Am. Chem. Soc.* **109**, 2837–2839 (1987).
- Raja, R. & Ratnasamy, P. Process for the oxidation of cyclohexane to a mixture of cyclohexanol and cyclohexanone. *US Patent* No. 5,767,320 (1998).
- Raja, R. & Ratnasamy, P. Selective oxidation with copper complexes incorporated in molecular sieves. *Stud. Surf. Sci. Catal.* **101**, 181–190 (1996).
- Maschmeyer, T. et al. Designing a solid catalyst for the selective low-temperature oxidation of cyclohexane to cyclohexanone. *Angew. Chem. Int. Edn* **36**, 1639–1642 (1997).
- Thomas, J. M. et al. On the nature of the active site in a CoAPO-18 solid acid catalyst. *Angew. Chem. Int. Edn* **33**, 1871–1873 (1994).
- Barrett, P. A., Sankar, G., Catlow, C. R. A. & Thomas, J. M. X-ray absorption spectroscopic study of Brønsted, Lewis and redox centres in cobalt-substituted aluminium phosphate catalysts. *J. Phys. Chem.* **100**, 8977–8985 (1996).
- Iton, L. E., Choi, I., Desjardins, J. A. & Maroni, V. A. Stabilisation of Co(III) in aluminophosphate molecular sieve frameworks. *Zeolites* **9**, 535–538 (1989).
- Lin, S. S. & Weng, H. S. Liquid-phase oxidation of cyclohexane over CoAPO-5: Synergism effect of co-reactant and solvent effect. *Appl. Catal.* **118**, 21–31 (1994).
- Krauschaar-Czarnetzki, B., Hoogervorst, W. G. M. & Stork, W. H. J. Oxidation of saturated hydrocarbons involving CoAPO molecular sieves as oxidants and as catalysts. *Stud. Surf. Sci. Catal.* **84**, 1869–1876 (1994).
- Tolman, C. A., Druliner, J. D., Nappa, M. J. & Herron, N. in *Activation and Functionalization of Alkanes* (ed. Hill, C. L.) 303–360 (Wiley, Chichester, 1989).
- Kerr, J. A. Bond dissociation energies by kinetic methods. *Chem. Rev.* **66**, 465–498 (1966).
- Chen, J. & Thomas, J. M. MAPO-18 (M = Mg, Zn, Co)—A new family of catalysts for the conversion of methanol to light olefins. *J. Chem. Soc. Chem. Commun.* 603–604 (1994).

- Thomas, J. M. The ineluctable need for in-situ methods of characterising solid catalysts as a prerequisite to engineering active sites. *Eur. J. Chem.* **3**, 1557–1562 (1997).
- Vanoppen, D. L., De Vos, D., Genet, M. J., Rouxhet, P. G. & Jacobs, P. A. Cobalt-containing molecular-sieves as catalysts for the low conversion autooxidation of pure cyclohexane. *Angew. Chem. Int. Edn* **34**, 560–563 (1995).
- Sankar, G., Raja, R. & Thomas, J. M. Redox solid catalysts for the selective oxidation of cyclohexane in air. *Cat. Lett.* **55**, 15–23 (1998).
- Thomas, J. M. & Thomas, W. J. *Introduction to the Principles of Heterogeneous Catalysis* 383–384 (Academic, London, 1967).
- Groves, J. T. Artificial enzymes—the importance of being selective. *Nature* **389**, 329–330 (1997).
- Ludwig, M. L., Metzger, A. L., Patridge, K. A. & Stallings, W. C. Manganese superoxide dismutase from *Thermus thermophilus*—a structural model refined at 1.8 Å resolution. *J. Mol. Biol.* **219**, 335–358 (1991).
- Barrett, P. A., Sankar, G., Jones, R. H., Catlow, C. R. A. & Thomas, J. M. Interaction of acetonitrile with cobalt-containing aluminophosphates: an x-ray absorption investigation. *J. Phys. Chem.* **101**, 9555–9562 (1997).
- Freeman, C. M., Catlow, C. R. A., Thomas, J. M. & Brode, S. Computing the location and energetics of organic molecules in microporous adsorbents and catalysts—hybrid approach applied to isomeric butenes in a model zeolite. *Chem. Phys. Lett.* **186**, 137–142 (1991).
- Morohashi, K., Sadano, H., Okada, Y. & Omura, T. Position specificity in *n*-hexane hydroxylation by two forms of cytochrome P-450 in rat liver microsomes. *J. Biochem.* **93**, 413–419 (1983).
- Hamberg, M., Samuelsson, B., Bjorkhem, I. & Danielsson, H. in *Molecular Mechanisms of Oxygen Activation* (ed. Hayaishi, O.) 24–52 (Academic, New York, 1974).
- Binsted, N. EXCURV 98 (CCLRC Daresbury Lab., 1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank D. W. Lewis for discussions. We also thank the UK EPSRC for a rolling grant (J.M.T.) and the Royal Commission for the Exhibition of 1851 for a research fellowship (R.R.). Molecular Simulation Inc. (MSI) is acknowledged for the provision of molecular modelling software.

Correspondence and requests for materials should be addressed to J.M.T.

Snowpack production of formaldehyde and its effect on the Arctic troposphere

Ann Louise Sumner* & Paul B. Shepson*†

Departments of * Chemistry and † Earth and Atmospheric Sciences, Purdue University, West Lafayette, Indiana 47907, USA

The oxidative capacity¹ of the atmosphere determines the lifetime and ultimate fate of atmospheric trace species. It is controlled by the presence of highly reactive radicals, particularly OH[•] formed as a result of ozone photolysis. The dramatic depletion of ozone in Arctic surface air during polar sunrise^{2,3} therefore offers an opportunity to improve our understanding of the processes controlling ozone abundance and hence the oxidative capacity of the atmosphere. Ozone destruction is catalysed by bromine atoms⁴ and is terminated once bromine reacts with formaldehyde to form relatively inert hydrogen bromide, but neither the activation of bromine nor the contribution of formaldehyde are fully understood. Particularly troubling is the failure of current models^{5–7} to simulate the high formaldehyde concentrations^{7–9} in Arctic surface air. Here we report measurements in Arctic snow and near-surface air, which suggest that photochemical production at the air–snow interface accounts for the discrepancy between observed and predicted formaldehyde concentrations. The strength of this source is comparable to that of the dominant formaldehyde source in the free troposphere (the reaction between OH[•] and methane) and implies that formaldehyde photolysis can be a dominant source of oxidizing free radicals in the lower polar troposphere. We expect that formaldehyde will also affect photochemistry at the snow surface to facilitate the release of bromine into the lower troposphere—the initial step in Arctic tropospheric ozone depletion.

We measured gas-phase formaldehyde (HCHO) at Canadian Forces Station Alert, Northwest Territories, Canada (82.5° N, 62.3° W) from 15 February to 26 April 1998 using a flow injection analysis system¹⁰. The study period ranges from days with 24 hours of darkness to days with 24 hours of daylight. During this period, we

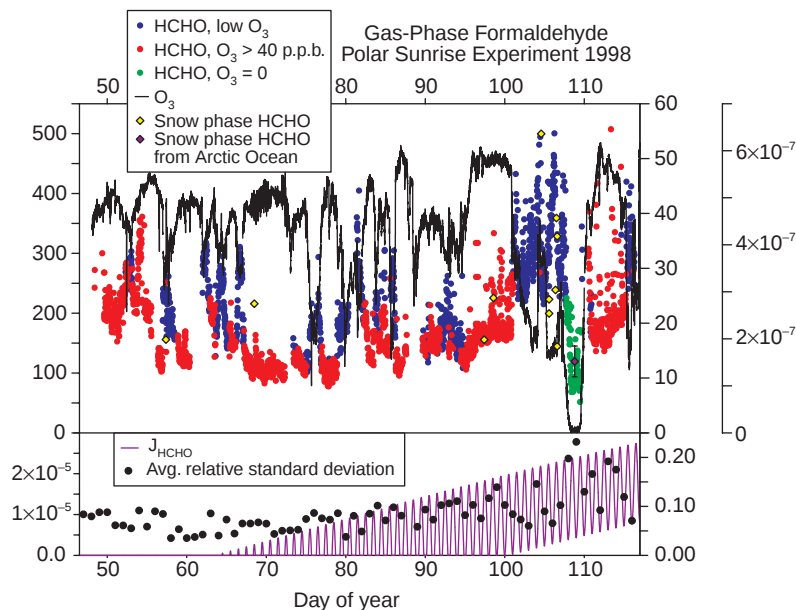


Figure 1 Results of gas-phase measurements. Shown are measurements of gas-phase HCHO (red, blue and green), snow-phase HCHO (yellow and purple diamonds), Ozone (black lines), J_{HCHO} (purple), and the average relative standard deviation for gas-phase formaldehyde (black dots). Formaldehyde measurements were segregated by the corresponding ozone concentration such that the values in red indicate high ozone (well-mixed southerly flow), blue values indicate partially-ozone-depleted air (some halogen atom processing), and green values indicate a completely-ozone-depleted air mass, representative of arctic MBL air. Gas-phase HCHO was measured with a continuous-flow injection analysis system¹⁰ that uses a Nafion diffusion membrane to extract ambient HCHO into water. This solution is reacted with NH_4^+ and 1,3-cyclohexanedione, in acid (the Hantzsch reaction), forming a rigid, planar, fluorescence product. The fluorescence signal at $\lambda = 465\text{ nm}$ resulting from irradiation at $\lambda = 390\text{ nm}$ is then detected. Sample injection was automated, on a 20-min sample cycle, with gas-

phase calibration samples injected every 2 h. The estimated accuracy, due to uncertainties in the calibration standard concentration, for the gas-phase measurements is $\pm 25\%$; the measurement precision is $\pm 2\%$. We also measured gas-phase HCHO in the snowpack interstitial air and analysed melted snow samples from the area (shown in yellow), and from the Arctic Ocean surface $\sim 65\text{ km}$ north of the Ellesmere Island coast (shown in purple). The snow samples, $\sim 50\text{ ml}$ in volume, were taken from the surface of the snowpack (upper $\sim 4\text{ cm}$) and stored in glass or Teflon bottles at local ambient temperature ($\sim 245\text{ K}$) and protected from light until analysis. The samples were melted $< 1\text{ h}$ before analysis, using the instrument described above, bypassing the Nafion membrane. Due to uncertainties in the snow sample stability, the snow-phase measurements carry $\pm \sim 40\%$ uncertainty, while the measurement precision is $\pm 11\%$. Standard HCHO solutions were used for calibration. During the measurement period, the average snow density was 0.32 g ml^{-1} .

also measured HCHO concentrations in snowpack interstitial air and in melted snow samples collected within the study area and from the Arctic Ocean surface $\sim 65\text{ km}$ to the north of the Ellesmere Island coast. The data we obtained are shown in the top panel of Fig. 1, together with gas-phase ozone concentrations at Alert (K. Anlauf, personal communication). The transition from darkness to sunlight began near day 75, with the 24-hour sunlight period starting around day 100. The bottom panel of Fig. 1 gives calculated values for the HCHO photodissociation rate coefficient (J_{HCHO}) and the 24-hour average of the running relative standard deviation of the HCHO measurements.

By differentiating between different ozone concentration regimes, the formaldehyde data are divided into three categories to indicate differences in the air mass origin and the dominant chemical processing before arrival at the measurement site. Specifically, when the concentration of O_3 is greater than 40 parts per billion (p.p.b., mole/mole) flow is generally from the south and represents a well mixed tropospheric O_3 "background" composition¹¹. When ozone is partially depleted, we believe that the air mass constitutes a mixture between O_3 -depleted Arctic marine boundary layer (MBL) air and undepleted "background" air. We distinguish this condition from the case when O_3 is completely removed (that is, Arctic MBL air) to examine the effect of O_3 as a reactant. The HCHO concentrations measured in the snow samples are also shown in Fig. 1, including the data for the snow samples taken over the Arctic Ocean on day 113. The latter data are positioned over day 109, when the ozone concentration at Alert dropped to essentially zero for a period of days. The air mass

sampled during this depletion event is likely to have been transported over the polar ice cap² and is thus representative of the air over the Arctic Ocean snow-sampling site.

The gas-phase HCHO concentrations measured ranged from 78 to 372 parts per trillion (p.p.t., mole/mole) for the dark period and 52 to 690 p.p.t. for the sunlit period, showing a strong inverse correlation with ozone, as long as O_3 was present to some extent. The range of concentrations and the inverse correlation with ozone are consistent with previous observations^{7,8}, but we also observe an HCHO minimum during the ozone depletion event, with values as low as 52 p.p.t. Although such low HCHO concentrations are expected in completely ozone-depleted air^{5,7}, they have not been previously observed, nor has the positive correlation with ozone in Arctic MBL air.

The relatively high concentrations of HCHO measured during the dark period can be explained by transport from anthropogenic source regions, as the HCHO lifetime is estimated to be ~ 160 days (ref. 8), with loss principally via dry deposition. In the presence of sunlight, photolysis decreases the lifetime to ~ 13 hours (refs 4, 7). Because the $\text{OH}^\cdot$ and NO_x levels in the Arctic are quite low^{7,12-14}, destruction of HCHO by photolysis is expected to be considerably faster than its production by known gas-phase chemistry (that is, the reaction of $\text{OH}^\cdot$ with hydrocarbons) in the sunlit period. Thus, the gas-phase concentrations would be expected to decrease as the solar elevation increases. As shown in Fig. 1 by the data in red, which we believe represents air not directly influenced by the ocean and gas-phase halogen chemistry, we observe the opposite: as the Sun rises, the formaldehyde concentrations increase in both concentration

and variability, suggesting that an additional local source of formaldehyde sustains the high concentrations of HCHO which we observe.

To investigate the role of the snowpack on formaldehyde, gas-phase HCHO vertical profiles were obtained. Figure 2 shows profiles measured on day 68, when there were ~6 hours of twilight, and under daylight conditions on days 97 and 106. As shown in Fig. 2, we observed that HCHO in the snowpack interstitial space increased in concentration with depth. The higher snowpack concentration implies that HCHO is produced in, and diffuses out of, the snowpack. In fact, Fuhrer *et al.*⁹ observed high hydrogen peroxide (H₂O₂) concentrations at Summit, Greenland, which they attributed to the photolysis of HCHO formed in snow. As we observe smaller HCHO concentrations during the dark period, when the HCHO lifetime is long, than during the sunlit period, HCHO production in the snowpack appears to occur through photochemical processes. (A light-independent HCHO formation mechanism would lead to a gradual build-up of HCHO during the dark period, which is in contrast with our observations.)

Using the HCHO concentrations measured on day 106, a snow depth of 53 cm and a MBL height of 600 m, we calculate that 77% of the total HCHO in the combined snowpack/MBL column resides in the snowpack. If we assume that HCHO is equilibrated between the air and snow phases, this equates to an equilibrium constant of $(8.1 \pm 3.5) \times 10^3$ molecules per cm³(snow)/molecules per cm³(air). This value agrees well with the equilibrium constants of 1.1×10^4 (ref. 15) and $(7.7\text{--}9.6) \times 10^3$ (ref. 6) inferred from formaldehyde measurements at Summit, Greenland, and implies that the snowpack source may control HCHO concentrations in the polar MBL. The snow/air equilibrium constants are ~60 times smaller than the equilibrium constant calculated for the HCHO equilibrium between the gas phase and liquid water at 245 K (ref. 16). This difference may reflect the fact that much of the snowpack HCHO is probably contained within the quasi-liquid water layer¹⁷ of the ice crystals, rather than being distributed throughout the bulk.

The magnitude of the HCHO production rate in the snowpack can be estimated by assuming that the gas-phase concentration of HCHO is in steady state, with a concentration of [HCHO]_{ss} = 150 p.p.t. (corresponding to air not affected by halogen chemistry), and that photolysis is the dominant process leading to the destruction of HCHO. The reactions that we have to consider are therefore:

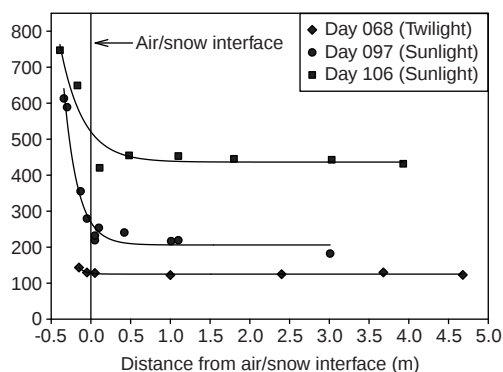
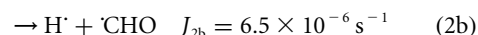
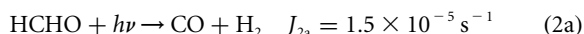


Figure 2 Vertical HCHO concentration profiles. To investigate the gas-snow equilibrium, HCHO vertical profiles were obtained by moving the heated air-sample inlet vertically from ~4 m above the snow surface, down through a 2-inch square hole extending to the snow/ice base. The inlet was positioned, secured in place, and allowed to equilibrate for at least 10 min before sampling at each height. The snow depth averaged 50 cm, with the deepest measurement made ~40 cm below the snow-air interface. The change in interstitial air concentration with depth into the snowpack clearly exceeds the measurement precision.



(The products of reaction (2b) react rapidly with molecular oxygen to form hydroperoxyl radicals (HO₂) and carbon monoxide: $\text{CHO} + \text{H} + 2\text{O}_2 \rightarrow 2\text{HO}_2 + \text{CO}$.) The two photodissociation rate coefficients (*J*) are those for clear-sky conditions at 82.5°N at noon on 13 April; *k*₁ represents the zero-order volumetric HCHO production rate resulting from its flux out of the snowpack. Under steady state conditions, $k_1 = (J_{2a} + J_{2b}) \times [\text{HCHO}]_{\text{ss}}$, leading to $k_1 = 8.1 \times 10^4$ molecules HCHO cm⁻³ s⁻¹. To understand the magnitude of this source we can compare it to the dominant source for HCHO production in the clean troposphere; that is, the reaction between the hydroxyl radical (OH[•]) and methane (CH₄). Using a global average [OH[•]] = 1×10^6 molecules cm⁻³ (ref. 18) and $k_{\text{OH}+\text{CH}_4} = 1.5 \times 10^{-15}$ molecules⁻¹ cm³ s⁻¹ (ref. 19, *T* = 245 K), we obtain a production rate of 8.1×10^4 molecules HCHO cm⁻³ s⁻¹. This value, which is an upper limit as we assume that the reaction always produces HCHO, is equal to our calculated production rate attributed to the snowpack source.

On the assumption that the HCHO produced at a volumetric rate of 8.1×10^4 molecules cm⁻³ s⁻¹ mixes throughout a 600-m-deep MBL, we can calculate a flux from the surface of $\sim 4.9 \times 10^9$ molecules HCHO cm⁻² s⁻¹. If this flux originates, as we expect, from a local (that is, spatially variable) surface source, it should lead to considerable variability in the gas-phase HCHO

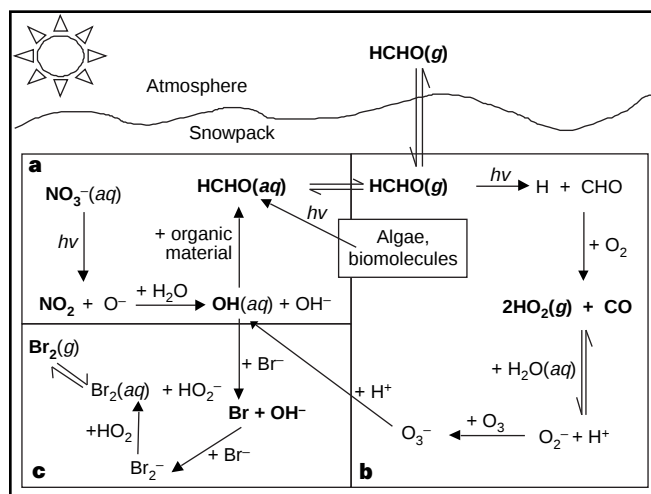


Figure 3 Possible formaldehyde production mechanisms and photochemical reaction sequences occurring at the snow-air interface. Direct sources for the formaldehyde observed may be emission by algae contained in polar snow and the photolysis of organic matter, including marine biomolecules (see text for further details). We also expect formaldehyde formation to occur as a result of photochemically activated radical chemistry as sunlight can penetrate ~10 cm into the snow⁶. Moreover, air exchange between the firn gas layer and the atmosphere occurs rapidly within the first metre, and the cold snow surface will act as a repository for low-volatility organic material. The photolysis of radical precursors within the quasi-liquid layer¹⁷ of the snow crystals, shown in **a** for nitrate (NO₃), may initiate radical chemistry with organic compounds present in the snowpack. These organic compounds are likely to include water-soluble organic ions such as acetate, lactate and oxalate anions, which have been identified in Arctic aerosols and snow²⁹, and humic acid from local sources. As illustrated in **b**, formaldehyde itself can contribute to snow-phase radical chemistry through its photolysis in snow pack interstitial space, resulting in the formation of OH radicals within the quasi-liquid layer of the snow crystals (see text for further details). The OH[•] produced may then participate in the release of photolytically active halogens from sea salt (**c**), thereby providing an indirect mechanism for formaldehyde to contribute to ozone destruction in the polar MBL.

concentration. The 24-hour running averages of the relative standard deviation of the HCHO measurements (Fig. 1, bottom panel) are calculated for each consecutive 80-minute time period, which is much shorter than the HCHO lifetime. As shown in the figure, the running averages, which give an indication of HCHO variability, increase significantly after sunrise, thus supporting the existence of a relatively strong local HCHO source activated by sunlight.

Figure 3 indicates possible sources of the observed HCHO flux leaving the snowpack. These include direct emission by algae, reflecting the observation that polar snow contains a range of different alga species²⁰, which may include some of the species known to produce formaldehyde²¹. As sunlight can penetrate snow to a depth of ~10 cm (ref. 6), the photolysis of organic material adsorbed to snow crystals or present in the quasi-liquid layer of the crystals¹⁷ may also give rise to HCHO production; this process is analogous to the photolysis of organic matter, including marine biomolecules, as a source of formaldehyde in Antarctic surface waters²². These two direct mechanisms are likely to occur along with photolytically activated radical chemistry, involving radical reaction with organic molecules present in the snowpack. For example, there is strong evidence that photolysable halogen atom precursors are produced in snow^{12,23,24}, with experimental studies indicating that the interaction between ozone and sea ice yields molecular bromine and chlorine^{25,26}. Similarly, OH[•], one of the most powerful oxidants, may be present in snow as a result of nitrate photolysis (Fig. 3a). This process is known to occur in ocean-surface²⁷ and natural²⁸ waters, and is likely to occur also in the quasi-liquid layer of snow crystals. Once the oxidizing free radicals have been photolytically produced, they will react with organic matter present in the snowpack²⁹ to form, among other products, formaldehyde (Fig. 3a).

Once formed, HCHO itself can be photolysed in the snowpack interstitial space to produce HO₂ radicals, which in turn can enter the quasi-liquid layer of the snow crystals and be converted into more reactive OH radicals³⁰ (Fig. 3b). This reaction sequence thus provides an autocatalytic mechanism for HCHO formation. It may also affect ice-core records of species susceptible to local chemical reactions within the snow, such as those of H₂O₂ (ref. 6), CO (ref. 31) and HCHO itself (ref. 15). The OH radicals originating from HCHO photolysis can, moreover, initiate the oxidation of halides to molecular halogens (Fig. 3c). HCHO may therefore be more strongly implicated in the release of reactive halogen species from the snowpack, and thus in the catalytic ozone destruction at polar sunrise, than previously assumed. In addition to this indirect effect on the chemistry of the MBL, the HCHO formed will also directly affect the Arctic troposphere, as exchange between the firn gas layer and the atmosphere occurs quickly within the first metre of the snowpack. Our measurements confirm that formaldehyde originating from the snowpack constitutes an important local HCHO source that can explain the discrepancy between model predictions^{5–7} and observed HCHO concentrations in Arctic surface air^{7–9}. Under conditions of low ambient humidity, when radical production through ozone photolysis is inefficient¹, HCHO photolysis is likely to constitute the most important radical source in the lower polar troposphere. □

Received 1 September; accepted 30 December 1998.

- Thompson, A. M. The oxidizing capacity of the Earth's atmosphere: probable past and future changes. *Science* **256**, 1157–1165 (1992).
- Bottenheim, J. W., Gallant, A. J. & Brice, K. A. Measurements of NO_y species and O₃ at 82°N latitude. *Geophys. Res. Lett.* **13**, 113–116 (1986).
- Oltmans, S. J. & Komhyr, W. D. Surface ozone distributions and variations from 1973–1984 measurements at the NOAA Geophysical Monitoring for Climate Change Baseline observatories. *J. Geophys. Res.* **D 91**, 5229–5236 (1986).
- Barrie, L. A., Bottenheim, J. W., Schnell, R. C., Crutzen, P. J. & Rasmussen, R. A. Ozone destruction and photochemical reactions at polar sunrise in the lower Arctic atmosphere. *Nature* **334**, 138–141 (1988).
- Sander, R., Vogt, R., Harris, G. W. & Crutzen, P. J. Modeling the chemistry of ozone, halogen compounds, and hydrocarbons in the arctic troposphere during spring. *Tellus B* **49**, 522–532 (1997).
- Nefel, A., Bales, R. C. & Jacob, D. J. in *Ice Core Studies of Global Biochemical Cycles* (ed. Delmas, R.) (NATO ASI Ser. I, Springer, New York, 1995).
- Shepson, P. B. et al. Sources and sinks of carbonyl compounds in the Arctic Ocean boundary layer: polar ice floe experiment. *J. Geophys. Res.* **101**, 21081–21089 (1996).

- De Serves, C. Gas phase formaldehyde and peroxide measurements in the Arctic atmosphere. *J. Geophys. Res.* **99**, 25391–25398 (1994).
- Fuhrer, K., Hutterli, M. & McConnell, J. R. in *Chemical Exchange between the Atmosphere and Polar Snow* (eds Wolff, E. & Bales, R.) (NATO ASI Ser. I, Springer, New York, 1996).
- Fan, Q. & Dasgupta, P. K. Continuous automated determination of atmospheric formaldehyde at the parts per trillion level. *Anal. Chem.* **66**, 551–556 (1994).
- Hopper, J. F. & Hart, W. R. Meteorological aspects of the 1992 Polar Sunrise Experiment. *J. Geophys. Res.* **99**, 25315–25328 (1994).
- McConnell, J. C. et al. Photochemical bromine production implicated in Arctic boundary layer ozone depletion. *Nature* **355**, 150–152 (1992).
- Honrath, R. E. & Jaffe, D. A. The seasonal cycle of nitrogen oxides in the Arctic troposphere at Barrow, Alaska. *J. Geophys. Res.* **97**, 20615–20630 (1992).
- Beine, H. J. et al. NO_x during ozone depletion events in the Arctic troposphere at Ny-Ålesund, Svalbard. *Tellus B* **49**, 556–565 (1997).
- Staffelbach, T., Nefel, A., Stauffer, B. & Jacob, D. A record of the atmospheric methane sink for formaldehyde in polar ice cores. *Nature* **349**, 603–605 (1991).
- Dong, S. & Dasgupta, P. K. Solubility of gaseous formaldehyde in liquid water and generation of trace standard formaldehyde. *Environ. Sci. Technol.* **20**, 637–640 (1986).
- Conklin, M. H. & Bales, R. C. SO₂ uptake on ice spheres: Liquid nature of the ice-air interface. *J. Geophys. Res.* **98**, 16851–16855 (1993).
- Prinn, R. G. et al. Atmospheric trends and lifetime of CH₃CCl₃ and global OH concentrations. *Science* **269**, 187–192 (1995).
- Vaghjiani, G. L. & Ravishankara, A. R. New measurement of the rate coefficient for the reaction of OH with methane. *Nature* **350**, 406–409 (1991).
- Muller, T., Bleiss, W., Martin, C. D., Rogaschewski, S. & Fuhr, G. Snow algae from northwest Svalbard: Their identification, distribution, pigment and nutrient content. *Polar Biol.* **20**, 14–32 (1998).
- Stewart, W. D. P. (ed.) *Algal Physiology and Biochemistry* (Univ. California Press, Berkeley, 1974).
- Qian, J.-G., Mopper, K. & Kieber, D. J. Photochemical production of the OH radical and formaldehyde in Antarctic waters. *Antarctic J. R.* **30**, 141–143 (1995).
- Impey, G. A., Shepson, P. B., Hastie, D. R. & Barrie, L. A. Measurements of photolysable chlorine and bromine during the Polar Sunrise Experiment 1995. *J. Geophys. Res.* **102**, 16005–16010 (1997).
- Impey, G. A. et al. Measurements of photolysable halogen compounds and bromine radicals during the Polar Sunrise Experiment 1997. *J. Atmos. Chem.* (in the press).
- Oum, K. W., Lakin, M. J., DeHaan, D. O., Brauers, T. & Finlayson-Pitts, B. J. Formation of molecular chlorine from the photolysis of ozone and aqueous sea-salt particles. *Science* **279**, 74–77 (1998).
- Oum, K. W., Lakin, M. Y. & Finlayson-Pitts, B. J. Bromine activation in the troposphere by the dark reaction of O₃ with seawater ice. *Geophys. Res. Lett.* **25**, 3923003926 (1998).
- Mopper, K. & Zhou, X. Hydroxyl radical photoproduction in the sea and its potential impact on marine processes. *Science* **250**, 661–664 (1990).
- Vaughan, P. P. & Blough, N. V. Photochemical formation of hydroxyl radical by constituents of natural waters. *Environ. Sci. Technol.* **32**, 2947–2953 (1998).
- Li, S.-M. & Winchester, J. W. Water soluble organic constituents in Arctic aerosols and snow pack. *Geophys. Res. Lett.* **20**, 45–48 (1993).
- Mozurkewich, M. Mechanism for the release of halogens from sea-salt particles by free radical reactions. *J. Geophys. Res.* **100**, 14199–142107 (1995).
- Haan, D. & Raynaud, D. Ice core record of CO variations during the last two millennia: atmospheric implications and chemical interactions within the Greenland ice. *Tellus B* **50**, 253–262 (1998).

Acknowledgements. We thank J. Bottenheim, L. Barrie and P. Dasgupta for discussions; K. Anlauf for use of the Alert ozone data; and J. Deary, A. Gallant, Atmospheric Environment Service, the Canadian Forces, and the Department of National Defense, Canada, for technical and logistic support. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to P.B.S. (e-mail: pshepson@purdue.edu).

Increased marine sediment suspension and fluxes following an earthquake

Robert Thunell*, Eric Tappa*, Ramon Varela†, Martin Llano†, Yrene Astor†, Frank Muller-Karger‡ & Richard Bohrer‡

* Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

† Estacion de Investigaciones Marinas de Margarita, Fundacion La Salle de Ciencias Naturales, Aptdo. 144, Porlamar 6301, Venezuela

‡ Department of Marine Science, University of South Florida, St Petersburg, Florida 33701, USA

Earthquakes are commonly cited as one possible triggering mechanism for turbidity flows—dense sediment–water plumes that can transport large volumes of sediment great distances down slope—in both marine and lacustrine settings^{1–6}. Heezen and Ewing¹ were the first to make such a suggestion, attributing breaks in a sea-floor telephone cable in the North Atlantic Ocean to turbidity flows generated by the 1929 Grand Banks earthquake. A number of workers have consequently used sedimentary turbidite records to reconstruct the earthquake histories of various regions^{2,7,8}. Here we present direct observations of a seismically

induced turbidity flow. Measurements of light scattering and sediment fluxes in the Cariaco basin indicate that the earthquake that occurred along the coast of northern Venezuela on 9 July 1997 resulted in considerable downslope displacement of sediments—probably $>10^5$ tonnes into the deep part of the basin. In such a seismically active region, this mechanism of sediment transport may be responsible for a significant component of the long-term sediment accumulation in the basin. Furthermore, this process may result in the sequestration in deep sea sediments of large amounts of carbon initially deposited at shallow depths.

In November 1995, a time-series observation campaign was initiated in the Cariaco basin to study the relationship between surface ocean processes and the downward flux of particulate material. This is a cooperative US–Venezuelan research effort and is referred to as CARIACO (for ‘carbon retention in a colored ocean’). Among other things, our study includes routine hydrographic monitoring, transmissometer measurements of light scattering by particles from the sea surface to the sea floor, and sediment trapping to estimate particle fluxes at a single site on the eastern side of the Cariaco basin ($10^{\circ}30'N$, $64^{\circ}40'W$) (Fig. 1). The sediment-trap mooring is placed in the deepest part of the basin ($\sim 1,400$ m) and contains four automated traps positioned at water depths of approximately 240 m, 420 m, 900 m, and 1,220 m (Fig. 2). Each trap contains 13 collection cups and is programmed to collect samples sequentially over two-week intervals.

On 9 July 1997, an earthquake of magnitude 6.8 occurred along the coast of Venezuela and produced a large zone of surface rupture (Fig. 1). This was the strongest earthquake to occur in this region in over 30 years, and resulted in extensive damage in the towns of Cariaco and Cumana. Minor damage also occurred offshore on Isla Margarita. The distance from our study site in Cariaco basin to the edge of the rupture zone is ~ 90 km.

Hydrographic monitoring was carried out at the mooring site on 8 July 1997, the day before the earthquake, as part of our routine monthly sampling effort. The transmissometer profile collected on this date shows uniformly low beam attenuation below 400 m (Fig. 2). Using a previously published empirical relationship for this particular model of transmissometer (Sea Tec 25-cm pathlength transmissometer)⁹, we estimated particle concentrations using beam attenuation derived from light transmission measurements. The low beam attenuation values below 400 m on 8 July are representative of particle concentrations of less than 0.1 mg l^{-1} . The increase in attenuation between approximately 250 and 300 m is associated with moderately high particle concentrations (up to 0.3 mg l^{-1}) at the interface between oxic and anoxic conditions. Previous studies have shown that this interface is marked by both high particle concentrations and intense alterations of these particles^{10,11}. The 8 July profile is typical of what had been collected at this site on a monthly basis over the previous 18 months. These

data indicate that normally there is not a zone of resuspended sediments just above the sea floor in the Cariaco basin. In contrast, most open ocean regions are characterized by a well developed, bottom nepheloid layer¹².

Prompted by the opportunity to observe the effects of this large earthquake, the study site was reoccupied on 16 July, one week after the event. The transmissometer profile collected on this date is characterized by a fairly constant and low level of beam attenuation from approximately 400 to 1,100 m depth, comparable to what was observed over this depth interval just before the earthquake (Fig. 2). However, below 1,100 m, beam attenuation increased significantly, indicating an increase in water-column particle concentrations (Fig. 2). Specifically, between 1,100 and 1,250 m, there is a systematic increase in beam attenuation that equates to an order-of-magnitude increase in particle concentrations, from approximately 0.1 to 1.0 mg l^{-1} . Below, 1,250 m, there is a further increase in beam attenuation with maximum particle concentrations reaching $\sim 1.2 \text{ mg l}^{-1}$ near the sea floor ($\sim 1,400$ m). The transmissometer profile from 16 July also shows a reduction in particle concentration at the oxic/anoxic interface (Fig. 2). We attribute this to internal waves generated by the earthquake that caused mixing across this interface and dispersion of the material.

Data collected during a subsequent cruise in mid-August 1997 show that water-column sediment concentrations below 1,100 m were still above the pre-earthquake level but substantially lower than those measured in mid-July (Fig. 2). By mid-September 1997, beam attenuation throughout the entire water column was low, indicating that particle concentrations were low at all depths. Based on this, it appears that virtually all of the material that was resuspended and displaced into the centre of the basin as a result of the 9 July 1997 earthquake had settled out of the water column within a two-month period.

The transmissometer data can be used to estimate how much resuspended material was carried into the eastern Cariaco basin by the seismically induced turbidity flow. Based on the 16 July 1997 record, we assume that the resuspended material was only deposited at depths greater than 1,100 m and that this beam attenuation profile is typical of what existed across the entire basin. Integrating this sediment concentration profile over the area of the basin deeper than 1,100 m yields an estimate of approximately 145×10^3 tonnes of displaced material that was transported into the basin.

As this transmissometer data was collected one week after the earthquake, the predicted sediment concentrations probably do not represent the maximum values and thus, the calculated total sediment deposition is an underestimation. In fact, an alternative hypothesis could be proposed in which most of the resuspended material transported into the basin (particularly coarse-grained material) settled out of the water column before the transmissometer cast from 16 July 1997. Although we do not have the data to

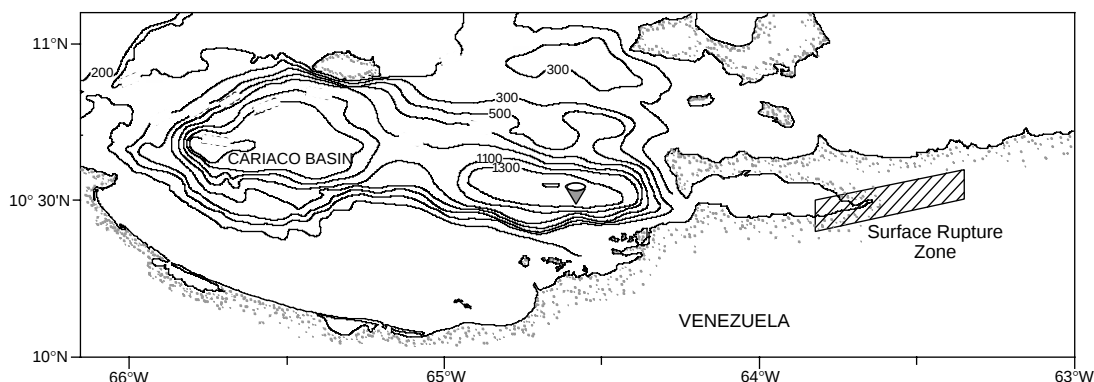


Figure 1 Bathymetric map of the study area. Locations of the sediment-trap mooring (cone) in Cariaco basin and the surface rupture zone of the 9 July 1997 earthquake are shown. Contours are in metres.

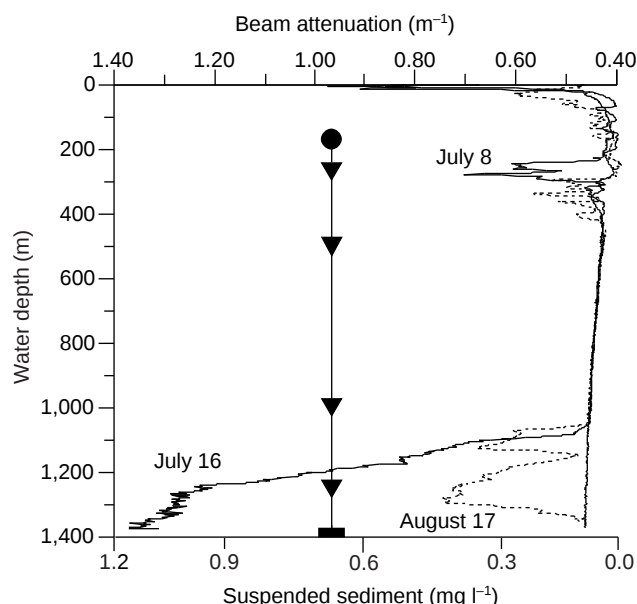


Figure 2 Transmissometer profiles collected at the sediment-trap mooring location. Beam attenuation values have been converted to sediment concentrations using the calibration of Spinrad⁸. The positions of the four sediment traps (triangles) on the mooring line are indicated. We note that only the deepest trap (~1,220 m) is within the zone of resuspended sediments.

evaluate this possibility, it would result in a significantly higher estimate of total displaced material.

The deepest sediment trap (1,220 m) on our mooring was positioned approximately 180 m above the sea floor and within the upper half of the zone of resuspended sediments, whereas the three shallower traps did not sample these sediments (Fig. 2). For the two-month period before the earthquake, the average sediment flux measured in the deepest trap was $0.35 \text{ g m}^{-2} \text{ d}^{-1}$ (Fig. 3). Moreover, for the two-week sampling period (19 June–2 July) just before the one during which the earthquake occurred, the sediment flux in the deepest trap was only $0.07 \text{ g m}^{-2} \text{ d}^{-1}$. In contrast, the average sediment flux in the 1,220 m trap for the 14-day sampling period (3–16 July) that included the earthquake was $4.33 \text{ g m}^{-2} \text{ d}^{-1}$. This is 25 times higher than the sediment flux measured in the 900 m trap for the same two-week period (Fig. 3). As this sampling period began at midnight on 2 July and the earthquake did not occur until 19:24 local time on 9 July, this average flux underestimates the true effect of the earthquake on the sediment flux. If we assume that the pre-earthquake portion of this sampling period had a sediment flux similar to the preceding sampling period ($0.07 \text{ g m}^{-2} \text{ d}^{-1}$), then the sediment flux for the remainder of the sampling period following the earthquake would be $8.38 \text{ g m}^{-2} \text{ d}^{-1}$. For the ensuing two sampling periods (17–30 July and 31 July–13 August) the sediment fluxes were significantly lower, but still above the pre-earthquake level at 0.67 and $0.26 \text{ g m}^{-2} \text{ d}^{-1}$, respectively. For the 14–27 August sampling period, sediment fluxes had decreased to $0.07 \text{ g m}^{-2} \text{ d}^{-1}$, as observed just before the earthquake.

Are these sediment fluxes consistent with what would be expected from the estimates of resuspended sediment? Based on the above sediment-trap measurements, the cumulative sediment flux in the deepest trap (1,220 m) was $\sim 74 \text{ g m}^{-2}$ for the period from 9 July to 13 August (the period during which there were resuspended material in the lower part of the water column). From the transmissometer profile of 16 July (Fig. 2), we estimate that the total sediment mass in a column of water over an area of $1 \times 1 \text{ m}$ between 1,100 and 1,220 m would be $\sim 60 \text{ g}$. Thus, the two measurements are within 20% of each other. Again, we expect this estimate of total

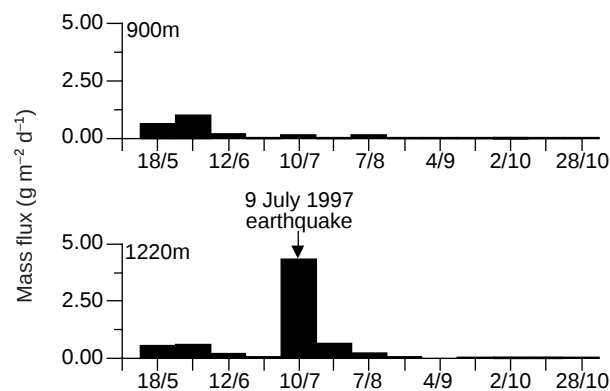


Figure 3 Total sediment fluxes measured at depths of 900 m and 1,220 m from May to November 1997. Each sampling period covers two weeks. The dates on the X-axis indicate (day/month) the mid-points for each sampling period. The 9 July 1997 earthquake occurred during the fifth sampling period which began on 3 July 1997.

resuspended sediments to be less than the measured cumulative sediment flux as the transmissometer reading was made one week after the earthquake, when sediment concentrations were probably not at their maximum.

The Cariaco basin and its surrounding area is a region of frequent earthquake activity; there were 92 earthquakes of magnitude 5.0 or greater in this region between 1900 and 1997¹³. Despite the large number of earthquakes, the sediment record in Cariaco basin for the past century contains only two distinctive turbidites. Based on ²¹⁰Pb dating, these turbidites have been attributed to the two largest earthquakes occurring in this region this century, the 1900 and 1929 earthquakes with magnitudes of 8.4 and 6.9, respectively⁶. These two earthquakes were located west of the July 1997 earthquake and were thus closer to Cariaco basin. In the centre of the eastern Cariaco basin, close to our mooring location, the turbidite formed in 1900 is $\sim 3 \text{ cm}$ thick and that formed in 1929 is $\sim 7 \text{ cm}$ thick⁶. We estimate that it would take total deposition rates of $6,000$ and $14,000 \text{ g m}^{-2}$ to produce the 3-cm- and 7-cm-thick turbidites, respectively. These values are significantly higher than both the sediment-trap flux measurements and the estimates of total resuspended sediments associated with the 1997 earthquake. In fact, based on our data, we predict that the turbidite produced by this most recent earthquake would be only 2–3 mm thick. Such millimetre-scale microturbidites are found dispersed throughout Cariaco basin sediment record⁶, and probably represent the typical sedimentary deposit associated with earthquakes in this region. Collectively, these microturbidites may account for a significant fraction of the total sediments accumulating in the deep Cariaco basin.

The measured carbon flux for the two-week period during which the earthquake occurred was $0.19 \text{ g C m}^{-2} \text{ d}^{-1}$, or 30 times higher than that measured for the preceding two-week period. Thus, these seismically induced turbidites could potentially serve as a mechanism for transferring, and rapidly burying in the deep ocean, significant amounts of carbon initially deposited in shallow regions. □

Received 8 July 1998; accepted 7 January 1999.

- Heezen, B. C. & Ewing, M. Turbidity currents and submarine slumps, and the 1929 Grand Banks earthquake. *Am. J. Sci.* **250**, 849–873 (1952).
- Inouchi, Y. *et al.* Turbidites as records of intense paleoearthquakes in Lake Biwa, Japan. *Sedim. Geol.* **104**, 117–125 (1996).
- Grantz, A. *et al.* Character, paleoenvironment, rate of accumulation, and evidence for seismic triggering of Holocene turbidites, Canada Abyssal Plain, Arctic Ocean. *Mar. Geol.* **133**, 51–73 (1996).
- Gorsline, D. S. Depositional events in Santa Monica Basin, California Borderland, over the past five centuries. *Sedim. Geol.* **104**, 73–88 (1996).
- Thomson, J. & Weaver, P. P. E. An AMS radiocarbon method to determine the emplacement time of recent deep-sea turbidites. *Sedim. Geol.* **89**, 1–7 (1994).

6. Huguen, K., Overpeck, J. T., Peterson, L. C. & Anderson, R. F. in *Palaeoclimatology and Palaeoceanography from Laminated Sediments* (ed. Kemp, A. E. S.) 171–183 (Spec. Publ. 116, Geol. Soc. London, 1996).
7. Adams, J. Paleoseismicity of the Cascadia subduction zone: evidence from turbidites off the Oregon–Washington margin. *Tectonics* **9**, 569–583 (1990).
8. Doig, R. 2300 yr history of seismicity from silting events in Lake Tadiussac, Quebec. *Geology* **18**, 820–823 (1990).
9. Spinrad, R. W. A calibration diagram of specific beam attenuation. *J. Geophys. Res.* **91**, 7761–7764 (1986).
10. Wakeham, S. G. & Ertel, J. R. Diagenesis of organic matter in suspended particles and sediments in the Cariaco Trench. *Org. Geochem.* **13**, 815–822 (1987).
11. Wakeham, S. G. Reduction of stenols to stanols in particulate matter at oxic–anoxic boundaries in sea water. *Nature* **342**, 787–790 (1989).
12. Biscaye, P. E. & Eittrheim, S. L. Suspended particulate loads and transports in the nepheloid layer of the abyssal Atlantic Ocean. *Mar. Geol.* **23**, 155–172 (1977).
13. US Geological Survey, National Earthquake Information Center (cited 25 May 1998) (<http://www.neic.cr.usgs.gov/>).

Acknowledgements. We thank the crew of the RV *Hermano Gines* for their assistance at sea, and W. Gardner, B. Kneller and R. Russo for comments that substantially improved the manuscript. This work was supported by the US NSF and the Venezuelan CONICIT.

Correspondence and requests for materials should be addressed to R.T. (e-mail: thunell@geol.sc.edu).

The limits of selection during maize domestication

Rong-Lin Wang*, Adrian Stec*, Jody Hey†, Lewis Lukens* & John Doebley*

* Department of Plant Biology, University of Minnesota, St Paul, Minnesota 55108, USA

† Department of Genetics, Rutgers University, Piscataway, New Jersey 08854-8082, USA

The domestication of all major crop plants occurred during a brief period in human history about 10,000 years ago¹. During this time, ancient agriculturalists selected seed of preferred forms and culled out seed of undesirable types to produce each subsequent generation. Consequently, favoured alleles at genes controlling traits of interest increased in frequency, ultimately reaching fixation. When selection is strong, domestication has the potential to drastically reduce genetic diversity in a crop. To understand the impact of selection during maize domestication, we examined nucleotide polymorphism in *teosinte branched1*, a gene involved in maize evolution². Here we show that the effects of selection were limited to the gene's regulatory region and cannot be detected in the protein-coding region. Although selection was apparently strong, high rates of recombination and a prolonged domestication period probably limited its effects. Our results help to explain why maize is such a variable crop. They also suggest that maize domestication required hundreds of years, and confirm previous evidence that maize was domesticated from Balsas teosinte of southwestern Mexico.

Several lines of evidence indicate that maize is a domesticated form of the wild Mexican grass teosinte (*Zea mays* ssp. *parviglumis* or ssp. *mexicana*)^{3–5}. Archaeological evidence places the time of maize domestication between 5,000 and 10,000 BP⁶. Despite the recent derivation of maize from teosinte, these plants differ profoundly in morphology⁵. One major difference is that teosinte typically has long branches with tassels at their tips whereas maize possesses short branches tipped by ears. Genetic analyses have identified *teosinte branched1* (*tb1*) as the gene that largely controls this difference⁷. Recent cloning of *tb1* (ref. 7) provides the first opportunity to examine the effects of selection on a 'domestication gene' and to infer from these effects the nature of the domestication process.

During development, *tb1* acts as a repressor of organ growth in those organs in which its messenger RNA accumulates. Consistent with this interpretation, plants carrying the maize allele accumulate more *tb1* mRNA in lateral-branch primordia and have shorter branches (that is, greater repression of branch elongation) than

plants carrying the teosinte allele, which accumulate less *tb1* mRNA and have longer branches⁷. This difference in message accumulation between the maize and teosinte alleles suggests that the evolutionary switch from teosinte to maize involved changes in the regulatory regions of *tb1*.

Domestication should strongly reduce sequence diversity at genes controlling traits of human interest. To test this expectation for *tb1*, we sampled a 2.9-kilobase (kb) region (Fig. 1) including most of the predicted transcriptional unit (TU) and 1.1 kb of the 5' non-transcribed region (NTR) from a diverse sample of maize and teosinte (Table 1). Two measures of genetic diversity were calculated: π , the expected heterozygosity per nucleotide site, and θ , an estimate of $4N_e\mu$, where N_e is the effective population size and μ the mutation rate per nucleotide⁸. Within the TU, maize possesses 39% of the diversity found in teosinte, which is not significantly lower than that (71%) seen for the neutral gene, *Adh1* (Table 2). However, within the NTR, maize possesses only 3% of the diversity found in teosinte. Thus, selection during domestication is associated with strongly reduced diversity in the NTR where regulatory sequences are typically found, but more modestly reduced diversity in the TU.

If *tb1* contributed to the morphological evolution of maize, then in the *tb1* phylogeny for maize and teosinte, maize sequences should form a single clade with only minor differentiation among them. Moreover, the type of teosinte most closely related to the direct ancestor of maize should be associated with the maize clade. In contrast, previous research with neutral genes not involved in maize evolution has shown that maize sequences for such genes are dispersed among multiple clades owing to the effects of lineage sorting^{9–13}. A phylogeny for the *tb1* TU fits the expectation of a neutral gene, with maize sequences falling into multiple clades (Fig. 2a). However, the phylogeny for the NTR shows all maize on a single, well-supported clade (I) (Fig. 2b), as predicted for a gene involved in maize evolution. There are five teosinte sequences

Table 1 List of taxa and collections sampled

Taxon	Sample	Collection	Origin
MAIZE	1L, 1P	BOV396	Bolivia
	2L, 2P	ECU969	Ecuador
	3L, 3P	GUA131	Guatemala
	4L	JAL78	Mexico
	5P	JAL44	Mexico
	6L	MEX7	Mexico
	7L	MOR17	Mexico
	8L	PI213778	North Dakota
	9L	PI218177	Arizona
	10L	SIN2	Mexico
	11L	YUC7B	Mexico
	12L, 12P	VEN604	Venezuela
	13L	W22	Wisconsin
PAR	14L, 15L	BK2	Chilpancingo, Guerrero
	16L, 17L	BK4	Palo Blanco, Guerrero
	18L	K71-4	Palo Blanco, Guerrero
	19L-A, B	B4	Teloloapan, Guerrero
	20P	BK6	Valle de Bravo, Mexico
	21P	IN1480	Jirosto, Jalisco
	22P	IC308	Tzitzio, Michoacan
	23P	BK1	Mazatlan, Guerrero
	24P	Z967	El Rodeo, Jalisco
MEX	25L	I28620	Texcoco, Mexico
	26L-A, B	B2438	Nobogame, Chihuahua
	27P	D625	Durango, Durango
	28P	D642	Tlamanalco, Mexico
	29P	D479	Texcoco, Mexico
	30P	P11066	Degollado, Jalisco
	31P	W45461	Panindicuar, Michoacan
	32P	J110	Toluca, Mexico
DIP	33L	G1120	Las Joyas, Jalisco

Taxa include *Zea mays* ssp. *mays* (MAIZE), ssp. *parviglumis* (PAR) and ssp. *mexicana* (MEX) and *Zea diploperennis* (DIP). Sequences isolated by λ cloning (L) and PCR (P). Samples for which we isolated and sequenced the 3' end of the gene are shown in bold. Source of the teosinte collections include G. Beadle (B), B. Benz (Z), T. Cochrane (C), J. Doebley (D), R. Guzman (G), H. Iltis (I), T. A. Kato (K), J. Kermicle (J), M. Nee (N), L. Puga (P) and H. G. Wilkes (W).

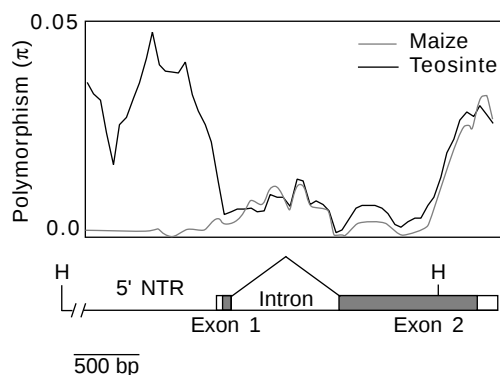


Figure 1 Predicted structure of *teosinte branched1* (ref. 25) and sliding-window analysis of polymorphism (π) in maize and teosinte. For the sliding-window analysis, π was calculated for segments of 300 bp at 50-bp intervals. Sequences used in the analysis were the subset of the λ -cloned sequences for which we isolated the 3' end by PCR (Table 1). The position of the *Hind*III (H) restriction endonuclease sites used in λ cloning are shown, as are the predicted exons (rectangles) and coding region (stippled).

tightly associated with the maize sequences and all belong to *ssp. parviglumis*. The teosinte (sample 16L) basal to clade I also belongs to *ssp. parviglumis*. This phylogeny and previous research¹⁴ provide compelling evidence that *ssp. parviglumis* (Balsas teosinte) is the progenitor of maize and suggest that maize arose in the Balsas river valley of southwestern Mexico, where this subspecies is native.

Although the phylogenies and the relative amount of nucleotide variation in maize suggest that selection has acted on *tb1*, we collected data on nucleotide polymorphism specifically to determine whether *tb1* has experienced a recent selective sweep. This determination was made using the HKA test¹⁵, in which the ratio of polymorphism within a species (maize) to divergence from an outgroup (*Z. diploperennis*) for *tb1* was compared with this ratio for neutral genes. A recent selective sweep in maize would be expected to reduce this ratio for *tb1* relative to neutral genes. The HKA test was not significant for the TU, indicating that there is no evidence for selection on the coding region (Table 3). However, the test was highly significant for the NTR, indicating that selection has strongly reduced variation here. Remarkably, the HKA test for the NTR was significant even if the TU was used as the control. This shows that the 'hitchhiking' effect was so small that it did not even affect the entire gene.

The relative impact of selection on the NTR and TU can be readily seen in a plot of polymorphism (π) for maize and teosinte along the length of *tb1* (Fig. 1). Throughout the NTR, π is substantially lower in maize than in teosinte, reflecting the impact of selection. At the boundary between the NTR and the TU, π for teosinte drops precipitously, as expected, because there is greater constraint on coding regions; however, π for maize rises until it is nearly equal to π for teosinte. Finally, approaching the stop codon at the 3' end of the gene, π rises steeply in both maize and teosinte, reflecting reduced constraint in this non-translated region. Figure 1 shows graphically how the impact of selection on polymorphism was narrowly focused on the NTR.

As further evidence that regulatory changes in the NTR rather

Table 2 Nucleotide polymorphism (π/θ) per bp ($\times 1,000$) in maize

Locus	MAIZE	MEX + PAR	PAR
<i>tb1</i> :NTR	0.47/0.93	33.08/35.52	28.68/32.57
<i>tb1</i> :TU	1.74/2.43	3.90/6.09	4.62/6.26
<i>Adh1</i> (ref. 9)	15.72/14.13	—	17.38/20.01

Taxa include *Zea mays* ssp. *mays* (MAIZE), *ssp. parviglumis* (PAR) and *ssp. mexicana* (MEX). These values are based on the 12 maize and 10 teosinte sequences that were cloned in λ (see Table 1).

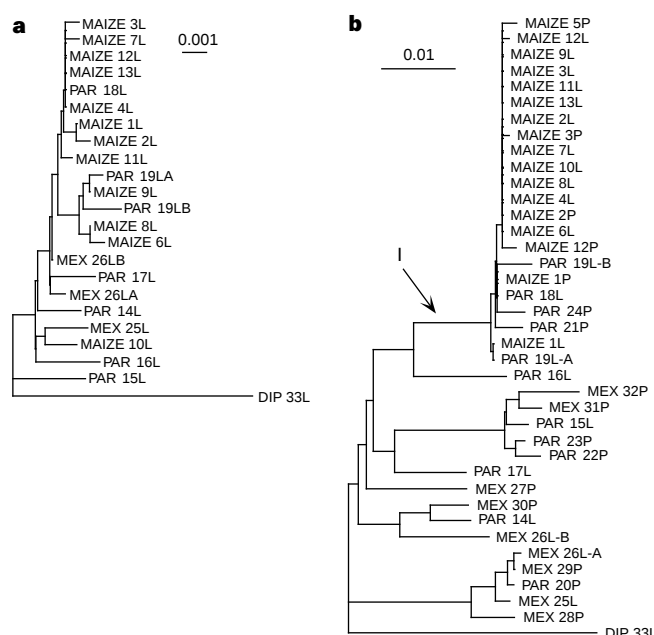


Figure 2 Neighbour-joining trees for *tb1* based on the 1,729-bp transcribed region (a) and the 1,143-bp 5' non-transcribed region (b). Taxa include maize, *ssp. parviglumis* (PAR), *ssp. mexicana* (MEX) and the outgroup *Z. diploperennis* (DIP). Sample numbers follow the taxon names. Scale bars indicate the number of substitutions per site using Kimura's 2-parameter distances. Clade I was supported in 200 of 200 bootstrap resamplings of the original data. An initial analysis of the 5' non-transcribed region using the 22 maize and teosinte λ -cloned sequences indicated that all maize alleles were derived from *ssp. parviglumis*. To confirm whether this result would be sustained with a larger sample, we isolated 16 additional sequences for a more comprehensive analysis (see Methods).

than changes in protein function were involved in maize evolution, we examined the predicted amino-acid sequence of our maize and teosinte sequences. Because our maize and teosinte λ clones did not include the 3' end of the TU unit, we isolated and sequenced an additional segment spanning the end of the λ clones to ~ 170 base pairs (bp) downstream of the stop codon (Table 1). Over the entire coding region, there are no fixed differences in the predicted amino-acid sequences between maize and teosinte.

Our analysis of nucleotide polymorphism in *tb1* provides compelling evidence that selection during maize domestication was aimed at the NTR, where regulatory elements are typically found. We had previously observed that the *tb1* mRNA for the maize allele accumulates at twice the level of that for the teosinte allele and proposed that changes in *tb1* regulation underlie maize evolution⁷. Combined evidence from polymorphism analysis and previous work on *tb1* mRNA levels are thus congruent, providing strong evidence that the short, ear-tipped lateral branches of maize evolved from the long, tassel-tipped branches of teosinte by human selection for novel regulatory elements in the NTR.

Although our data implicate selection on regulatory sequences during maize evolution, we found no fixed differences between

Table 3 HKA test of neutrality at *tb1* in maize

Test locus	<i>tb1</i> NTR	<i>tb1</i> TU	<i>tb1</i> NTR	<i>adh1</i> (ref. 9)	<i>adh2</i> (ref. 11)
Control loci	<i>adh1</i> , <i>adh2</i>	<i>adh1</i> , <i>adh2</i>	<i>tb1</i> TU		
Polymorphism	0.93	2.43	0.93	14.13	25.8
Divergence	52.55	12.73	52.55	21.25	23.6
Ratio	0.018	0.19	0.018	0.66	1.09
χ^2	13.58	2.70	8.24		
P	0.001	0.26	0.004		

Polymorphism (θ) and divergence (average pairwise differences between maize and *Z. diploperennis*) are reported per base pair ($\times 1,000$).

maize and teosinte within the 1.1 kb of NTR that we analysed. In fact, some maize and teosinte sequences (for example, maize 1P and *parviglumis* 18L) are nearly identical within this region, differing by only 1 bp in the length of a poly(A) track. This may indicate either that the selected site lies further upstream or that the differences between maize and teosinte are complex and depend on a group of polymorphisms rather than a single site¹⁶. Recombination between an upstream selected site and the region we sequenced could explain why maize is not fully separated from teosinte in the phylogeny (Fig. 2b).

Selection intensities during domestication are expected to be high because crop evolution involves dramatic changes in morphology within a short time. Under directional selection, the selection coefficient (s) is measured as the difference in relative fitness of the most fit and least fit genotypes, where fitness is the contribution of a genotype to the next generation. For example, $s = 0.01$ would indicate that 100 maize alleles would be passed to the next generation for every 99 teosinte alleles. A rough estimate of s requires a knowledge of the recombination rate (c , crossovers per bp per generation) and the distance (d) in bp from the selected site over which there has been a substantial reduction in nucleotide variation¹⁷:

$$d = 0.01s/c$$

For maize, recombination rates have been empirically measured for several genes, giving a mean value for c of $\sim 4 \times 10^{-7}$ (refs 18–21). Two observations allow a preliminary estimate of d : first, the substantial reduction in nucleotide variation is restricted to the NTR or promoter and does not extend into the TU, and, second, plant promoters are normally 2 kb or less in size (see Methods). Thus, the selected site is likely to be less than 2 kb from the TU and must be at least 1.1 kb from the TU since it does not appear to lie in the 1.1 kb of NTR that we sequenced. Using these values, s is estimated to be between 0.04 and 0.08. This estimate can be refined in the future by obtaining direct estimates of c in *tb1* and identifying the precise position of the selected site.

When s is known, one can estimate the time (T_f) required to bring the maize allele to fixation²². We assume that the initial frequency of the maize allele was $1/2N$, where N is the population size, and that gene action was additive². We considered two population sizes during the time of selection: 1,000, which assumes teosinte was grown like a horticultural crop in gardens, and 100,000, which assumes it was grown like an agriculture crop but still over a limited geographical area. We assume values for s of 0.04 and 0.08 (see above). For these values, T_f ranges from 315 to 1,023 years. Thus, the morphological evolution of maize as controlled by *tb1* could have been rapid, over just several hundred years.

We were surprised that maize remained polymorphic for *tb1*, even within the NTR. To assess whether the observed level of variation at *tb1* in maize is consistent with previous estimates of mutation and recombination rates, population sizes and the time of maize domestication, we carried out coalescent simulations²³. The simulations included a selective sweep modelled on a range of estimates for T_s (time since the selective sweep) and s (Table 4). To measure the effect of a selective sweep so that it reflects the present

context of not knowing the actual site of selection, we measured the longest segment with zero, one or two polymorphic sites. The simulated mean values of these lengths are remarkably close to the observed data. Thus, even for genes under strong selection, domestication need not remove all variation. The ability of maize to remain polymorphic at *tb1* probably reflects high recombination rates over the hundreds of years required to bring the maize allele to fixation such that there was substantial recombination between the allele that initially harboured the selected site and other alleles in the population. By this means, considerable polymorphism was maintained in the coding region in the face of strong selection on the NTR.

Population-genetic analysis of domestication genes can provide a new view of the processes that sculpted the formation of crop species. For *tb1*, such analysis indicates that ancient agriculturalists exerted a strong selective force on *tb1* that has drastically reduced polymorphism in its regulatory region but not in its coding region. This observation is consistent with previous evidence that alterations in the regulation of *tb1* brought about the change from teosinte to maize plant architecture. We also infer that it took at least several hundred years to bring the maize allele of *tb1* to fixation. Finally, these analyses indicate that Balsas teosinte is the ancestor of maize, since all maize alleles sampled show a close and statistically robust phylogenetic association with this teosinte. □

Methods

Gene isolation. All sequences for population-genetic analysis were cloned into λ -ZAP (Stratagene) as *Hind*III fragments (Fig. 1). Isolation of the 3' end of the gene for the sliding-window analysis (Fig. 1) and determination of the complete amino-acid sequence were accomplished by the PCR reaction (primers: TAGTTCATCGTCACACAGCC and CAATAACGCACACC AGGTCC). PCR was performed using PCR Supermix (Life Technologies) with one step of 4 min at 95 °C followed by 30 cycles of 1 min at 95 °C, 1 min at 60 °C and 3 min at 72 °C followed by 10 min at 72 °C. PCR products were cloned using the TOPO TA cloning kit (Invitrogen). Additional sequences of the NTR for phylogenetic analysis (Fig. 2) were isolated by PCR (primers: GCTATTGGCTACAAGTGACC and GGATAATGTGCACCAGGTGT). All sequences were deposited in GenBank (accession nos AF131649 to AF131705).

Statistics. Calculation of the range of reasonable values for s requires an estimate of the position of the selected site relative to the TU. The selected site is expected to lie within the NTR region in which regulatory sequences occur. Such regions typically extend 2 kb or less upstream of the TU in plants. For example, average gene density in *Arabidopsis* is one gene for every 4.8 kb²⁴. With an average gene being about 2.5 kb long, this leaves about 2 kb for flanking regulatory sequences. Moreover, many reports in the literature reveal that 5' regulatory sequences are usually within 1 kb of the transcription start site. For the coalescent simulations, the middle of the 4,000-bp sampled chromosomes was set as the point of a selective sweep. For the moderate values of s modelled, the use of deterministic allele frequency change will closely follow a stochastic selective sweep¹⁷. The population mutation rate (θ) was set to 0.0262 per bp, which is the estimated value from teosinte in the *tb1* NTR. The effective population size was set to the value (700,000) estimated for maize at *Adh1*, based on estimates of θ and μ for *Adh1* (ref. 9). The recombination rate (c) was set to 4×10^{-7} as described in the text. Sample size was 12, the same as the number of maize λ clones, and 200 runs were performed.

Received 27 November 1998; accepted 1 February 1999.

Table 4 Results from coalescent simulations of the effect of the selective sweep during maize domestication on polymorphism in *tb1*

T_s	s	Mean lengths (bp)			
		0 PS	1 PS	2 PS	
5,000	0.04	682.80	796.20	875.60	Pred.
10,000	0.04	566.00	725.80	830.40	Pred.
5,000	0.08	1,068.40	1,233.20	1,329.00	Pred.
10,000	0.08	892.00	1,133.00	1,321.00	Pred.
–	–	440	760	1,035	Obs.

T_s is the time in generations since the end of the selective sweep, and s is the selection coefficient. The predicted (Pred.) and observed (Obs.) mean lengths of segments with zero, one and two polymorphic sites (PS) are shown.

- Harlan, J. *Crops and Man* (Am. Soc. Agron., Madison, WI, 1992).
- Doebley, J., Stec, A. & Gustus, C. *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**, 333–346 (1995).
- Beadle, G. Teosinte and the origin of maize. *J. Hered.* **30**, 245–247 (1939).
- Galinat, W. The origin of maize as shown by key morphological traits of its ancestor, teosinte. *Maydica* **28**, 121–138 (1983).
- Ilitis, H. From teosinte to maize: the catastrophic sexual transmutation. *Science* **222**, 886–894 (1983).
- Smith, B. *The Emergence of Agriculture* (Freeman, New York, 1995).
- Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
- Watterson, G. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 188–193 (1975).
- Eyre-Walker, A., Gaut, R., Hilton, H., Feldman, D. & Gaut, B. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl Acad. Sci. USA* **95**, 4441–4446 (1998).
- Buckler, E. & Holtsford, T. Zea systematics: ribosomal ITS evidence. *Mol. Biol. Evol.* **13**, 612–622 (1996).

11. Goloubinoff, P., Paabo, S. & Wilson, A. Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc. Natl Acad. Sci. USA* **90**, 1997–2001 (1993).
12. Hanson, M. *et al.* Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* **143**, 1395–1407 (1996).
13. Hilton, H. & Gaut, B. Speciation and domestication in maize and its wild relatives. Evidence from the *glabulin-1* gene. *Genetics* **150**, 863–872 (1998).
14. Doebley, J., Goodman, M. & Stuber, C. Isoenzymatic variation in *Zea* (Gramineae). *Syst. Bot.* **9**, 203–218 (1984).
15. Hudson, R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
16. Stam, L. F. & Laurie, C. C. Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559–1564 (1996).
17. Kaplan, N., Hudson, R. & Langley, C. The “hitchhiking effect” revisited. *Genetics* **123**, 887–899 (1989).
18. Okagaki, R. & Weil, C. Analysis of recombination sites within the maize *waxy* locus. *Genetics* **147**, 815–821 (1997).
19. Patterson, G., Kubo, K., Shroyer, T. & Chandler, V. Sequences required for paramutation of the maize *b* gene map to a region containing the promoter and upstream sequences. *Genetics* **140**, 1389–1406 (1995).
20. Dooner, H. & Martinez-Ferez, I. Recombination occurs uniformly within the *bronze* gene, a meiotic recombination hotspot in the maize genome. *Plant Cell* **9**, 1633–1646 (1997).
21. Xu, X., Hsia, A., Zhang, L., Nikolau, B. & Schnable, P. Meiotic recombination break points resolve at high rates at the 5' end of a maize coding sequence. *Plant Cell* **7**, 2151–2161 (1995).
22. Kimura, M. & Ohta, T. The average number of generation until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771 (1969).
23. Hudson, R. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**, 183–201 (1983).
24. Bevan, M. *et al.* Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391**, 485–488 (1998).
25. Doebley, J. & Stec, A. The structure of *teosinte branched1*: a progress report. *Maize Genet. Coop. News* **1**, 73 (1998).

Acknowledgements. We thank E. Buckler, B. Gaut and J. Wendel for comments. This research was supported by the NSF and the Plant Molecular Genetics Institute of the University of Minnesota.

Correspondence and requests for materials should be addressed to J.D. (e-mail: doebley@tc.umn.edu).

Gaze direction controls response gain in primary visual-cortex neurons

Yves Trotter & Simona Celebrini

Centre de Recherche Cerveau et Cognition, Faculté de Médecine de Rangueil, Université Paul Sabatier, 133, route de Narbonne, 31062 Toulouse Cedex, France

To localize objects in space, the brain needs to combine information about the position of the stimulus on the retinae with information about the location of the eyes in their orbits. Interaction between these two types of information occurs in several cortical areas^{1–12}, but the role of the primary visual cortex (area V1) in this process has remained unclear. Here we show that, for half the cells recorded in area V1 of behaving monkeys, the classically described visual responses are strongly modulated by gaze direction. Specifically, we find that selectivity for horizontal retinal disparity—the difference in the position of a stimulus on each retina which relates to relative object distance—and for stimulus orientation may be present at a given gaze direction, but be absent or poorly expressed at another direction. Shifts in preferred disparity also occurred in several neurons. These neural changes were most often present at the beginning of the visual response, suggesting a feedforward gain control by eye position signals. Cortical neural processes for encoding information about the three-dimensional position of a stimulus in space therefore start as early as area V1.

Area V1 is the first cortical area where orientation and horizontal retinal disparity are encoded^{13–15}. Here, cells have oriented receptive fields that may occupy disparate locations on both retinae. Most of these cells have their activity (visual and/or spontaneous) modulated by the viewing distance in the straight-ahead sagittal direction^{16,17}. But do such modulations also occur as a function of the direction of gaze? This would imply that V1 cells would be more dedicated to certain volumes of visual space, in which case changing the direction of gaze should affect some or all of the visual proper-

ties encoded in the primary visual cortex, such as horizontal retinal disparity and orientation selectivity.

We obtained data from 142 neurons in two monkeys that were trained to fixate a target at three different positions in the fronto-parallel plane (Fig. 1a). For studies of both disparity and orientation, changes in gaze direction produced significant changes in neuronal activity in 54% ($n = 67$) of cells tested for disparity and 50% ($n = 104$) tested for orientation. The main effect was a significant change in the evoked firing rate (gain) in 72% of cells studied for disparity and in 85% studied for orientation. Shifts in preferred disparity angle were observed in 17% of cells; the remainder showed inconclusive changes in the tuning curves. Three examples of the gain effect on disparity coding are shown in Fig. 2. The cell shown in Fig. 2a is disparity selective with the preferred response in the plane of fixation (0°) when the monkey fixates in the centre of the screen or on the left, but shows a significant drop in the level of visual response, close to the spontaneous activity level, when the monkey fixates on the right. The cell shown in Fig. 2b exhibits significant progressive increase in the evoked firing rate in the plane of fixation (tuned 0°) from the left to the right direction of gaze. That shown in Fig. 2c displays a shift in preferred disparity angle: it has a preferred disparity angle in the plane of fixation (0°) for a gaze directed to the left, but shifts its peak just behind that plane (centred on 0.2°) for the right direction, with an intermediate step for the straight-ahead direction.

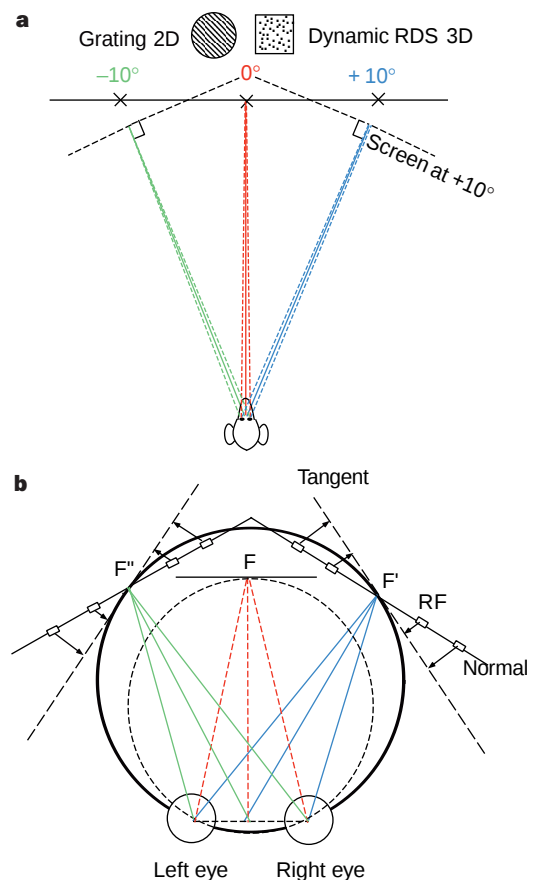


Figure 1 Experimental set-up. **a**, Dynamic random dot stereograms (RDS) and square-wave gratings were flashed on a video monitor screen subtending 42° or 32° of visual angle at three directions of gaze (straight ahead, 0° ; left, -10° ; and right, $+10^\circ$) in the fronto-parallel plane. For the left and right directions, the video monitor was rotated by 10° to maintain geometrical configurations with the viewing distance kept constant at 50 cm. Continuous lines of view represent the binocular axis. **b**, Vieth-Müller circles passing through both eyes and through fixation point for the three directions of gaze (F, F' and F'') (adapted from ref. 19).

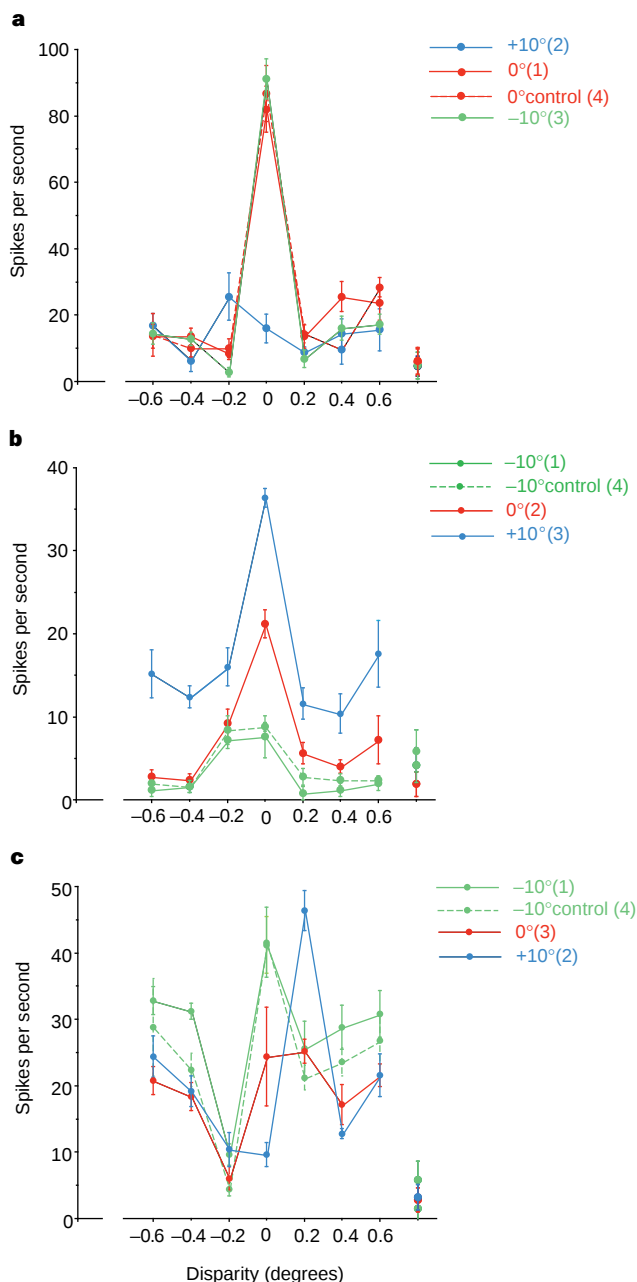


Figure 2 Retinal disparity tuning curves obtained in three individual neurons at three directions of gaze (-10° , green; 0° , red; $+10^\circ$, blue). **a–c**, The three individual neurons. Numbers in parentheses indicate the temporal sequence of recordings. For each cell, the first condition (curve 1) was repeated (control, dotted curve 4) at the end of the session of tests. The level of spontaneous activity is indicated on the right of the curves. Vertical bars show standard errors of the mean (five trials).

Three examples of the gain effect on orientation are shown in Fig. 3. The cell shown in Fig. 3a is visually responsive with a preferred orientation of 45° for the straight-ahead direction; a significant decrease in visual response occurs for the left direction and there is an almost total loss of visual response for the right direction. The cell shown in Fig. 3b is responsive when the monkey fixates on the left, but the level of visual response drops significantly for the other gaze directions. Finally, the cell in Fig. 3c shows a clear visual response (22.5°) for the left, but not for the right, direction of gaze.

We tested 29 cells with both types of stimuli. Among the 38% of cells that showed an effect of the two properties, only one showed a modulation for orientation but not for disparity; in all other cases, the effects were congruent for a common gaze direction.

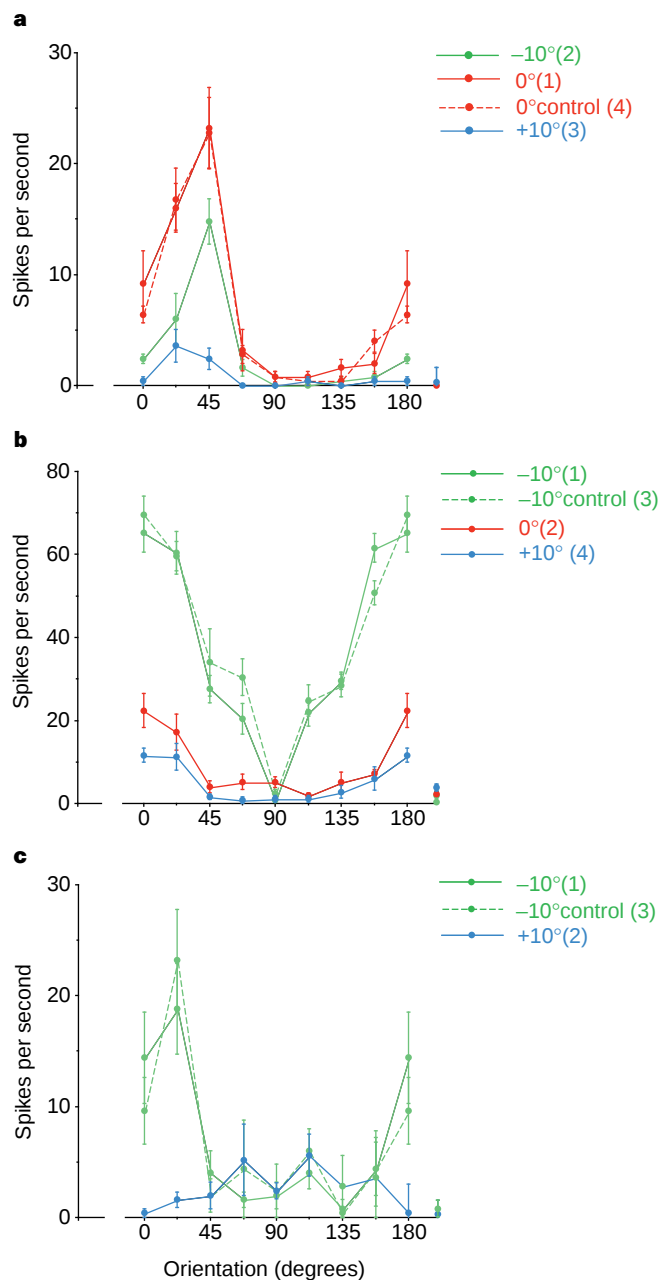


Figure 3 Effects of changing gaze direction on the responses of three individual neurons to oriented stimuli. Otherwise as Fig. 2; the three neurons are shown in **a–c**.

We quantified differences in response magnitude between the gaze directions for the two properties. The distributions of the modulation index (Fig. 4) are similar for disparity and orientation studies: modulation index, mean 0.48 ± 0.05 for disparity and 0.53 ± 0.04 for orientation (analysis of variance, $P < 0.0005$ between distributions gain effect/no effect for both properties). From the distributions of the modulation index, it follows that about 50% of cells showing an effect have a ratio of more than two for the visual activity evoked for the 'best' over the 'worst' direction of gaze, a proportion similar to that described in the parietal cortex¹.

For both disparity and orientation properties, modulations of the amplitude of the neural discharge occurred for the three directions of gaze, with no preference for either the contralateral or the ipsilateral field of view, and occurred equally for cells recorded in supra- and infragranular layers. Effects were independent of the preferred orientations or disparity angles encoded by cells. The gaze

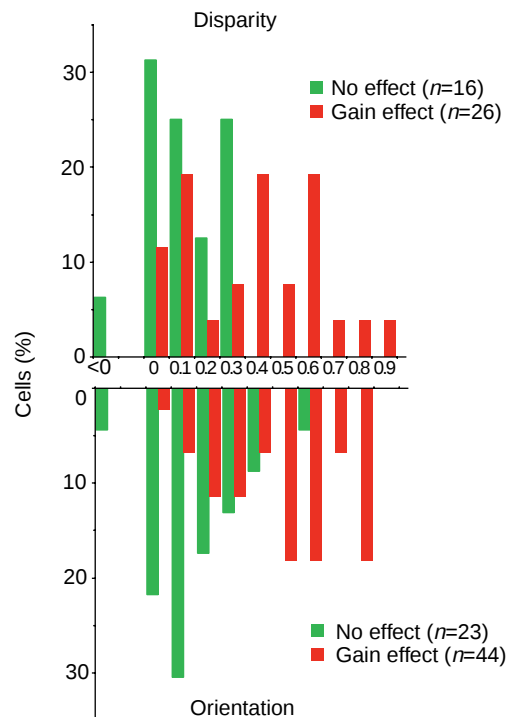


Figure 4 Distributions of the modulation index. Data are for horizontal disparity (top) and stimulus orientation (bottom) for cells showing a gaze direction effect (red) and cells that do not (green); *n*, number of cells.

direction also affected visually responsive cells that were non-stimulus selective, constituting 21% of disparity- and 9% of orientation-tested cells, in a similar way. Spontaneous activity was found to be modulated in only 11% of cases, with no correlation with the modulation on visual response.

Variations in the neural response are not due to effects such as fatigue or adaptation because controls of activity stability were performed for 90% of cells by repeating the first block of recordings after the last one to ensure that the tunings and levels of visual response remained similar. Turning the screen for the left and right directions of gaze, in order to maintain constant binocular distances of fixation, limits size deformation of the images on the retinae and so limits vertical retinal disparity¹⁸.

For a gaze directed in the centre of the screen (F in Fig. 1b; symmetrical convergence), the normal to the binocular axis and the tangent to the Vieth–Müller circle, which is used as a reference for stereoscopic judgments¹⁹, are superimposed. For a gaze directed on the left (F') or on the right (F'') (asymmetrical convergence), the normal to the direction of gaze is rotated away from the tangent, which should induce a change in binocular correspondence. Therefore, for the normal surface to yield correct stereoscopic spatial perception, a compensation process must take place^{19,20}. The subset of neurons that changes their preferred disparity angle may be the neural substrate that allows compensation for this shift in depth. For the example in Fig. 2c, the cell was recorded in the right hemisphere with the receptive field located in the left hemifield of view at 3° of retinal eccentricity. Any visual stimulus presented in its receptive field will appear 0.1° nearer than it should be for the fixation on the right, or 0.1° farther for a fixation on the left. So the shift of preferred disparity observed is a direct way of compensating for this misleading depth perception of the stimulus (as seen in five out of six cells).

The question arises if the effects of changing gaze direction are due to vertical disparity and/or oculomotor signals. Vertical disparity in our study is smaller than 1 min of arc, too small to account

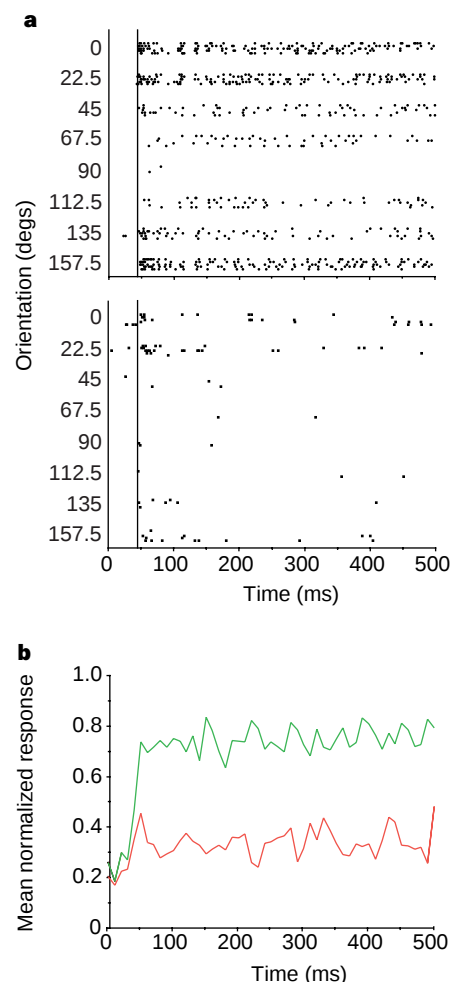


Figure 5 Time course of visual responses. **a**, Raster displays for an orientation-selective cell whose visual response was tested at -10° (top) and $+10^\circ$ (bottom). Vertical bars indicate the beginning of the visual responses. **b**, Mean time course of visual responses of the population of cells with a modulation index of more than 0.3 (17 cells for disparity and 35 cells for orientation) were pooled together as there was no difference in their mean time course). The responses of cells to the preferred stimulus for the gaze direction giving the higher activity (green) and that giving the lower activity (red) are normalized and averaged for the 52 cells by 10-ms time bins over the 500 ms of the visual response.

for the strong effects shown here. Moreover, six cells were tested monocularly for orientation selectivity, excluding vertical disparity, three of which showed clear modulations of neural activity as a function of gaze direction. This supports the idea that an eye-position signal is involved in the neural modulation process, at least for cells responding to oriented gratings. Other factors, such as contextual environment, attention or binocular fixation instability, are unlikely to be responsible for the modulations, as the monkeys were placed in total darkness, and the potential degree of attention required to fuse the target binocularly was similar for the three directions of gaze under the binocular control of eye movements. In addition, the variability of the visual response was similar for the three fixation locations.

This gain process, first described in parietal areas^{1,2} and interpreted as forming a distributed representation of space in head-centred coordinates^{2,21}, now appears as a common functional rule of the dorsal visual pathway. So perhaps our observations from area V1 are the consequence of feedback influences from higher integrated cortical areas? If this were the case, these influences on the visual response in area V1 would be reflected by a delay after the onset response in the pattern of the visual discharge, as has been shown for

contextual modulations outside the receptive field in area V1 of the behaving monkey²².

Two examples of raster displays of an orientation-selective cell in two conditions of fixation, left/right, are shown in Fig. 5a, where it can be seen clearly that the response is decreased (bottom raster) at the very beginning of the visual response. Indeed, the curves in Fig. 5b show that, on average, the difference in activity is present at the beginning of the visual response and remains constant throughout. This suggests that feedforward interactions may be involved in the modulation of visual activity by the gaze angle in area V1, a neural gating that could be mediated from the lateral geniculate nucleus, where eye-position signals have been shown to influence the visual activity^{23–25}.

Finally, area V1 appears to be an integrated cortical area in which attentional and contextual influences^{26–28} may take place in addition to vergence angle-related signals^{16,17} and, as we have shown, gaze direction signals. We propose that stimulus selectivity in area V1 is optimally expressed within restricted volumes of space. □

Methods

General. All experimental protocols, including care, surgery and training of animals, were performed according to the Public Health Service policy on the use of laboratory animals. Two monkeys (*Macaca mulatta*) were placed in complete darkness, with their head fixed, and trained to fixate on a small bright target (12 min of arc) on a video screen. Eye position was monitored using scleral search coils implanted in both eyes, and the monkeys maintained binocular fixation for random periods of 1–2 s, and were rewarded by a drop of fruit juice or water. All trials with binocular fixations shifted outside an angular window of $\pm 1^\circ$ were rejected. Receptive fields were located using a computer-controlled track-ball. Extracellular recordings were performed using insulated tungsten microelectrodes in area V1 within 4° of the foveal projection. Visual stimuli were flashed binocularly for 500 ms, and in six cases monocularly, centred on the receptive field and presented five times randomly interleaved for all disparity and orientation angles tested. Spike activity was collected 300 ms before the appearance of the fixation target (spontaneous activity) and 500 ms after the visual stimulus onset (evoked activity). Records were taken for three gaze directions in 75% of cells and two gaze directions in 25% of cells.

Visual stimulation. Three-dimensional stereoscopic stimulation was performed using dynamic random dot stereograms generated through ferroelectric stereo glasses (60 frames per s per eye), figure size $6^\circ \times 6^\circ$, dot density 20%, dot size 3.5 min of arc. Horizontal disparity (-0.6° to $+0.6^\circ$ by 0.2° steps) was introduced between each dot so the entire figure appeared in front (negative values), behind (positive values) or in the plane (0°) of the fixation target. Orientation selectivity (two-dimensional) was tested with square-wave gratings, flashed for 500 ms in a circular window (6°) with a spatial frequency of 2 cycles deg^{-1} in 8 steps of 22.5° angles for 180° .

Data analysis and controls. Disparity and orientation tuning curves were assessed for each direction of gaze using the same stimuli in identical retinotopic positions. Two-way analysis of variance was used to test for significant effects ($P < 0.05$) of the stimulus and of the direction of gaze on mean firing rates. For 90% of cells, the first tested direction was repeated at the end and, if the activity was significantly different ($P < 0.05$), together with visual inspection, the data were discarded. The modulation of visual activity (gain effect) was quantified for each cell by: $1 - (\text{min-SA})/(\text{max-SA})$, where min is the mean activity taken at the peak of the tuning for the gaze direction that evokes the lower activity, max is the mean maximum response for the gaze direction giving the larger response, and SA is the spontaneous activity. An index of 0.5 indicates a ratio of two between max and min. The approximate laminar location of cells (supra- or infragranular) was determined by combining physiological criteria of layer-4 location (noticeable by its high neuronal activity) and depth of recordings.

Received 22 December 1998; accepted 25 January 1999.

- Andersen, R. A. & Mountcastle, V. M. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **3**, 532–548 (1983).
- Andersen, R. A., Essick, G. K. & Siegel, R. M. Encoding of spatial location by posterior parietal neurons. *Science* **230**, 456–458 (1985).
- Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W. & Fogassi, L. Eye position effects on visual, memory and saccade-related activity in areas LIP and 7A of macaque. *J. Neurosci.* **10**, 1176–1196 (1990).

- Squatraro, S. & Maioli, M. G. Gaze field properties of eye position neurones in areas MST and 7a of the macaque monkey. *Vis. Neurosci.* **13**, 385–398 (1996).
- Galletti, C. & Battaglini, P. P. Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J. Neurosci.* **9**, 1112–1125 (1989).
- Galletti, C., Battaglini, P. P. & Fattori, P. Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *Eur. J. Neurosci.* **7**, 2486–2501 (1995).
- Weyand, T. G. & Malpeli, J. G. Responses of neurons in primary visual cortex are modulated by eye position. *J. Neurophysiol.* **69**, 2258–2260 (1993).
- Guo, K. & Li, C. Y. Eye position-dependent activation of neurones in striate cortex of macaque. *NeuroReport* **8**, 1405–1409 (1997).
- Newsome, W. T., Wurtz, R. H. & Komatsu, H. Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *J. Neurophysiol.* **60**, 604–644 (1988).
- Duhamel, J. R., Bremmer, F., BenHamed, S. & Graf, W. Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature* **389**, 845–848 (1998).
- Boussaoud, D., Barth, T. M. & Wise, S. P. Effect of gaze on apparent visual responses of monkey frontal cortex neurons. *Exp. Brain Res.* **93**, 423–434 (1993).
- Graziano, M. S. A., Hu, X. T. & Gross, C. G. Visuospatial properties of ventral premotor cortex. *J. Neurophysiol.* **77**, 2268–2292 (1997).
- Hubel, T. & Wiesel, T. N. Receptor fields and functional architecture of monkey striate cortex. *J. Physiol. (Lond.)* **195**, 215–243 (1968).
- Barlow, H. B., Blakemore, C. D. & Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *J. Physiol. (Lond.)* **193**, 327–342 (1967).
- Poggio, G. F. & Fischer, B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys. *J. Neurophysiol.* **40**, 1392–1407 (1977).
- Trotter, Y., Celebrini, S., Stricane, B., Thorpe, S. & Imbert, M. Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science* **257**, 1279–1281 (1992).
- Trotter, Y., Celebrini, S., Stricane, B., Thorpe, S. & Imbert, M. Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *J. Neurophysiol.* **76**, 2872–2885 (1996).
- Ogle, K. N. in *Spatial Localization Through Binocular Vision* (ed. Davson, H.) 271–324 (Academic, New York, 1962).
- Ogle, K. N. in *The Problem of the Horopter* (ed. Davson, H.) 325–348 (Academic, New York, 1962).
- Morrison, L. C. Stereoscopic localization with the eyes asymmetrically converged. *Am. J. Optom. Physiol. Optics* **54**, 556–566 (1977).
- Zipser, D. & Andersen, R. A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331**, 679–684 (1988).
- Zipser, K., Lamme, V. A. F. & Schiller, P. H. Contextual modulation in primary visual cortex. *J. Neurosci.* **16**, 7376–7389 (1996).
- Donaldson, I. M. L. & Dixon, R. A. Excitation of units in the lateral geniculate and contiguous nuclei of the cat by stretch of extrinsic ocular muscles. *Exp. Brain Res.* **38**, 245–255 (1980).
- Molotchnikoff, S. & Casanova, C. Reactions of the geniculate cells to extraocular proprioceptive activation in rabbits. *J. Neurosci. Res.* **14**, 105–115 (1985).
- Lal, R. & Friedlander, M. J. Gating of retinal transmission by afferent eye position and movement signals. *Science* **243**, 93–96 (1989).
- Motter, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* **70**, 909–919 (1993).
- Gilbert, C. D. Adult cortical dynamics. *Physiol. Rev.* **78**, 467–485 (1998).
- Lamme, V. A. F., Super, H. & Spekreijse, H. Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* **8**, 529–535 (1998).

Acknowledgements. We thank J. Bullier, Y. Frégnac and S. Thorpe for criticism of the manuscript; K. Britten for advice on various logistics; and M. Imbert for continuous support. This work was supported by the Centre National de la Recherche Scientifique.

Correspondence and requests for materials should be addressed to Y.T. (e-mail: trotter@cerco.ups-tlse.fr).

brinker is a target of Dpp in *Drosophila* that negatively regulates Dpp-dependent genes

Maki Minami*, Noriyuki Kinoshita†, Yuko Kamoshida*, Hiromu Tanimoto* & Tetsuya Tabata*

* Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-0032, Japan

† Department of Neurophysiology, Tokyo Metropolitan Institute for Neuroscience, 2-6 Musashidai, Fuchu-shi, Tokyo 183-8526, Japan

Growth and patterning of the *Drosophila* wing is controlled in part by the long-range organizing activities of the Decapentaplegic protein (Dpp)^{1–4}. Dpp is synthesized by cells that line the anterior side of the anterior/posterior compartment border of the wing imaginal disc. From this source, Dpp is thought to generate a concentration gradient that patterns both anterior and posterior compartments. Among the gene targets that it regulates are *optomotor blind* (*omb*)⁵, *spalt* (*sal*)⁶, and *daughters against dpp* (*dad*)⁷. We report here the molecular cloning of *brinker* (*brk*), and show that *brk* expression is repressed by *dpp*. *brk* encodes, a protein that negatively regulates Dpp-dependent genes. Expression

of *brk* in *Xenopus* embryos indicates that *brk* can also repress the targets of a vertebrate homologue of Dpp, bone morphogenetic protein 4 (BMP-4). The evolutionary conservation of Brk function underscores the importance of its negative role in proportioning Dpp activity.

Two types of gene have been identified as targets of Dpp regulation (Fig. 1): *omb* and *sal*, which encode transcription factors that are made in response to Dpp signalling, execute aspects of the Dpp-dependent programme of growth and patterning^{5,6}; and *dad*, which has a different role. Although expression of *dad* is also activated by Dpp, it downregulates Dpp signalling in the cells in which it is expressed^{7,8}, presumably serving to limit the extent of Dpp activity. The gene we now describe, *brinker* (*brk*) is also a Dpp target that negatively regulates Dpp signalling.

brk was identified by characterizing two strains with P-element insertions that generate expression patterns complementary to that of *dad* (Fig. 1e–h). Both strains (no. 38 and 3sB) have a P-element insertion in polytene cytological location 7B, where *brk*^{9,10} has been genetically mapped (S. Roth and E. Wieschaus, personal communication). We discovered that the original *brk* allele (a gift from S. Roth) fails to complement lethal alleles of strain 38 that were made either by excising (38-3) or remobilizing (38-20) the P element at 7B; we therefore refer to the gene at the strain 38 and 3sB insertion sites as *brk*. As the expression pattern of *lacZ* in strain 38 suggests that Dpp represses *brk* expression, we examined β -galactosidase activity in strain-38 wing discs in which a UAS-*dpp*

transgene is transcribed in a ring around a wing pouch under the control of the 30A-Gal4 driver¹¹ (Fig. 1i). These discs grew to be abnormally large, and β -galactosidase expression was absent from its normal domain of strain 38 expression (Fig. 1j). To confirm that *brk* is regulated by Dpp, β -galactosidase expression was examined in strain 38 wing discs that carried *Mad* mutant clones (*Mad* encodes an essential transcription factor that transduces the Dpp signalling¹²): β -galactosidase was autonomously upregulated in *Mad* mutant clones (Fig. 1k, l), whereas overexpression of *Mad* autonomously repressed β -galactosidase expression (data not shown). These results indicate that Dpp directly controls *brk* expression.

To isolate the *brk* gene, genomic DNA surrounding the strain-38 insertion was cloned by plasmid rescue. A genomic fragment containing transcribed DNA was identified by *in situ* hybridization and northern blot analysis and this fragment was used to isolate homologous complementary DNAs. cDNA e7-4 contains an open reading frame encoding a 77.5K, 704-residue protein. Although the deduced amino-acid sequence contains no known structural motif, it shares weak homology with the *Drosophila* transcription factor Doublesex¹³, suggesting the Brk protein may be a transcription factor (Fig. 2). The enhancer trap line of strain 38 has a P-element insertion 8 kilobases (kb) upstream for this transcription unit. In *brk*³⁸⁻²⁰, which is allelic to *brk* and was made by remobilization of the P element in strain 38, the P element is inserted in the transcription unit at a position corresponding to 250 nucleotides downstream of

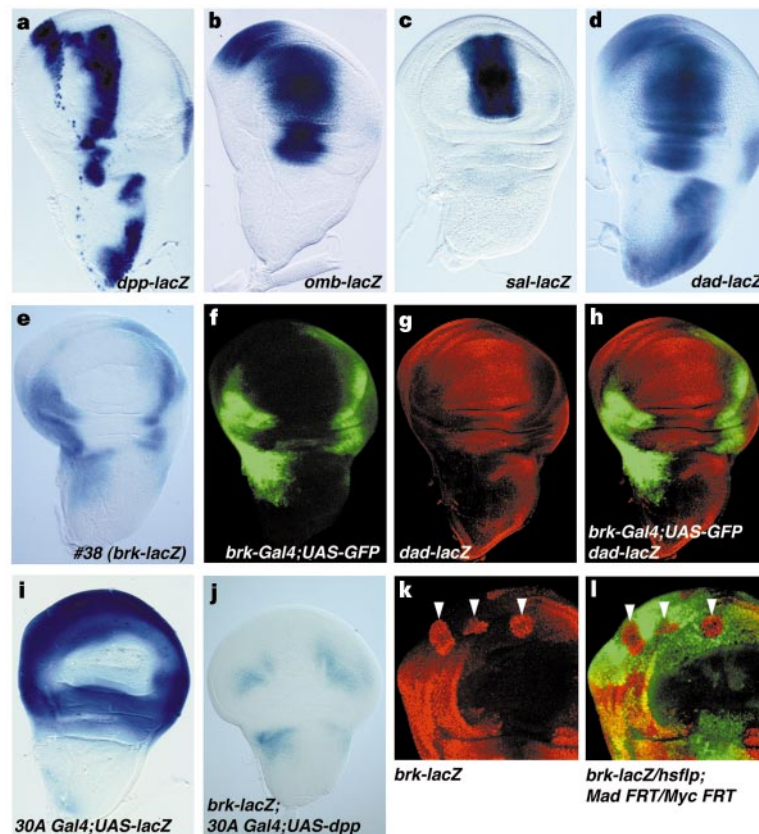


Figure 1 Expression of *brk* is negatively controlled by *dpp* in the wing imaginal disc. **a–e**, Wild-type expression patterns of *dpp* revealed by the enhancer trap *P1552* (**a**), *omb* revealed by the enhancer trap *omb*^{P1} (**b**), *sal* revealed by the enhancer trap *sal*^{JA405.1M2} (**c**), *dad* revealed by the enhancer trap *P1883* (**d**), and *brk* revealed by the enhancer trap #38 (**e**). **f–h**, Expression of the Gal4 enhancer trap line 3sB-Gal4 (F-A. Ramirez-Weber and T. Kornberg, unpublished) that carries an insertion at the *brk* locus and is complementary to that of *dad*. Discs from 3sB-Gal4/+;UAS-GFP/+;P1883/+ were stained against β -galactosidase. Expression of 3sB-Gal4 was monitored by green fluorescent protein (GFP) (**f**, green) and that of

dad by immunostaining against β -galactosidase (**g**; red). **h**, Superimposed images of **f** and **g**. **i–l**, *brk* expression, monitored by staining for β -galactosidase activity in *brk*^{*38}, is repressed by the Dpp signal. **i**, UAS-*lacZ* expression driven by 30A-Gal4. **j**, *brk* expression is repressed where UAS-*dpp* is driven by 30A-Gal4. **k**, **l**, *brk*-*lacZ* (red) is autonomously expressed in clones of cells mutant for *Mad* (arrowheads), identified by the absence of the Myc tag (green in superimposed image **l**). Images in **k** and **l** are shown at 1.7 times the magnification of **a–j**. All discs shown in this and subsequent figures are wing discs of third instar larvae. Posterior is to the right and ventral is up.

the 5' end of e7-4 cDNA. As re-excision of the P element restores Brk function in this line, we conclude that this cDNA is encoded by the *brk* gene.

To investigate the role of *brk*, we analysed mutant clones in wings and found that *brk* mutant clones caused outgrowth of wing tissue (Fig. 3b; compare to the wild type in Fig. 3a), a phenotype that is similar but not identical to the phenotype caused by ectopic *dpp* expression² (Fig. 3c). Both genes caused outgrowths, but the consequences of ectopic *dpp* expression were non-autonomous, whereas the effect of *brk* was autonomous in the mutant cells. We confirmed this by examining the expression of the Dpp-target genes *omb*, *sal* and *dad* in *brk*-mutant clones: we found that all three genes were expressed autonomously in mutant clones located within the normal domain of *brk* expression (Fig. 3d–l). The normal domain of *sal* expression is narrower than that of *omb* or *dad* and is nested in the domain where *brk* expression is very low (Fig. 1); *brk*-mutant cells adjacent to the normal domain of *sal* expression ectopically upregulate *sal* expression, indicating that the *brk* expression gradient is functional—it represses *omb* expression at high levels and *sal* expression at low levels. Clones within the domain of high *brk* expression were larger than those elsewhere, probably because of overexpression of Dpp targets in these cells.

We next examined *brk* expression in *brk*³⁸⁻²⁰ mutant clones by monitoring β -galactosidase expression (Fig. 3m–o). *brk*³⁸⁻²⁰ has an inserted P element in the *brk* transcription unit and lacks *brk* function. The patterns of *brk-lacZ* expression were as normal in mutant clones: not detectable in clones near the A/P compartment border and high in the peripheral clones, suggesting that *brk* expression is not regulated by Brk itself. Large clones straddling the region where *brk* expression is high had *brk-lacZ* expression that gradually declined towards the anterior–posterior (A/P) compartment border, as seen in wild-type discs. The relative position of the medial border at which *brk-lacZ* expression became undetectable in *brk* clones and in wild-type tissue was indistinguishable (Fig. 3m), indicating that *brk* expression is regulated by Dpp signalling in the absence of *brk* and that if *brk* is a component of Dpp signalling, it does not function upstream of *Mad*. If *brk* did function upstream of

Mad, *brk* mutations would affect graded *brk* expression because *brk* expression is regulated by Dpp signalling that is transduced by *Mad* (Fig. 1k, l), which they do not: *brk* may function parallel to, or downstream of, *Mad* when regulating Dpp target genes (Fig. 3p).

We investigated whether Brk could antagonize signalling by BMP-4, a vertebrate homologue of Dpp, because components of the Dpp/BMP signalling pathway are highly conserved between vertebrates and invertebrates¹⁴. In *Xenopus* embryos, BMP-4 specifies ventral mesodermal and epidermal fates. Blockade of BMP-4 signalling during gastrulation dorsalizes mesodermal tissues and neuralizes explanted ectodermal cells. When *brk* messenger RNA was injected into both blastomeres of two-cell embryos, the injected embryos showed a typical dorsalized phenotype, with a short trunk and a large cement gland, yielding an average dorsoanterior index (DAI; ref. 15) of 1.2 ($n = 16$) (Fig. 4a). To test directly whether Brk can dorsalize mesodermal tissue, we injected *brk* mRNA into two ventral blastomeres of four-cell embryos, isolated ventral marginal zone (VMZ) explants from early gastrulae, and cultured the explants in isolation until sibling embryos reached stage 30. VMZ explants from injected embryos expressed the muscle actin gene and formed immunoreactive muscle (Fig. 4b), indicating that ectopically expressed Brk can dorsalize ventral mesoderm. To test whether Brk can antagonize BMP function, we co-injected BMP-4 and *brk* mRNAs into embryos (Fig. 4a). Injection of BMP-4 mRNA alone caused a loss of anterior and dorsal structures (DAI = 0.9; $n = 22$), whereas co-injection of *brk* mRNA rescued the ventralized phenotype (DAI = 3.5; $n = 20$), indicating that Brk can antagonize BMP-4 function. In ectodermal explants (animal caps) from *brk*-injected embryos, neural-specific (*NCAM*) and cement gland-specific (*XAG*) gene expression was induced, whereas dorsal mesodermal (muscle actin) and panmesodermal (*Xbra*; data not shown) gene expression was not (Fig. 4c). This suggests that Brk, like known BMP antagonists¹⁶, can directly induce neural tissue in the absence of dorsal mesoderm. We next tested whether Brk could antagonize *Mad*. *Drosophila* *Mad* can function in the signalling pathway of BMP-4 in *Xenopus* embryos, and injection of *Mad* mRNA into the embryo causes animal caps to express a ventral mesodermal marker,

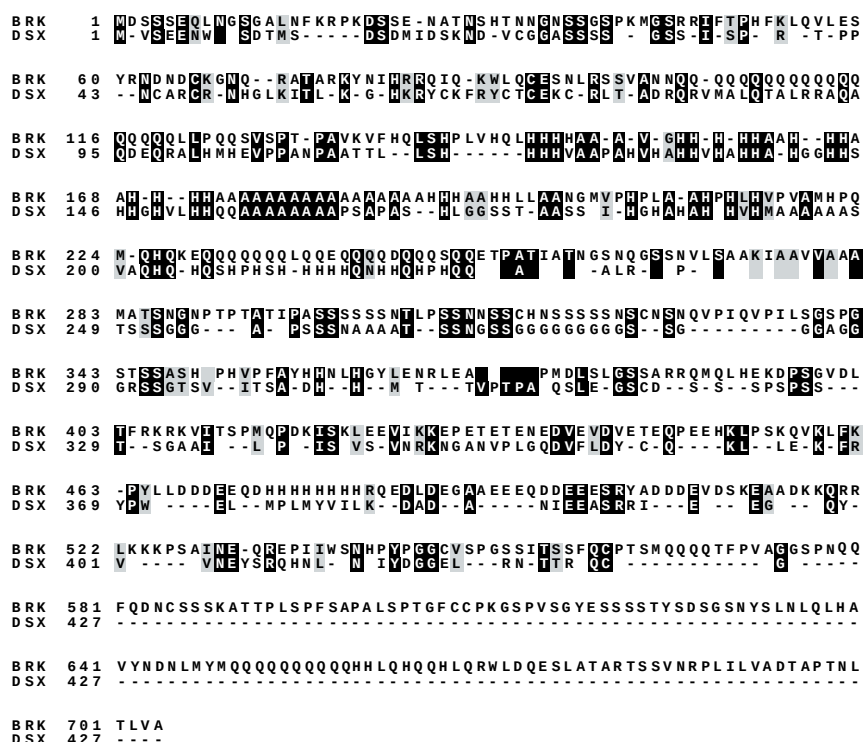


Figure 2 Deduced amino-acid sequence of Brk and alignment with *Drosophila* Dsx. Identical residues are highlighted against a black background and conserved residues are boxed in grey.

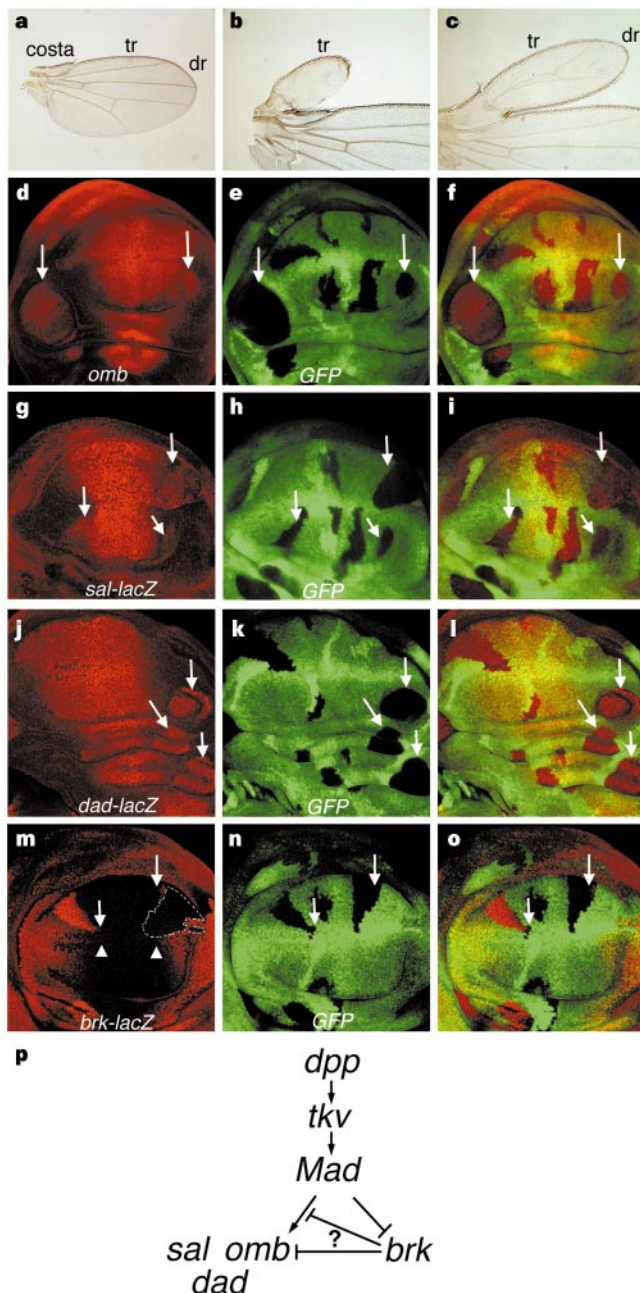


Figure 3 Phenotypes and gene expression patterns in somatic *brk* clones. **a–c**, Wings of wild type (**a**), with a *brk*-mutant clone marked by yellow (**b**), and with a *dpp*-expressing clone (**c**). Triple-row (tr) and double-row bristles (dr) and costa are indicated. Images in **b** and **c** are shown at twice the magnification used for **a**. *omb* (**d–f**), *sal*, (**g–i**), and *dad* (**j–l**) expression in *brk* clones. *omb* (**d**; red), *sal-lacZ* (**g**; red), and *dad-lacZ* (**j**; red) are autonomously expressed in clones of cells mutant for *brk* (arrows), identified by the absence of GFP (**e**, **h**, **k**; green). **f**, **i**, **l**, Superimposed images of **d** and **e** are shown in **f**, of **g** and **h** in **i**, and of **j** and **k** in **l**. **m–o**, *brk* expression pattern does not change in *brk* clones. The relative position of the medial border at which *brk-lacZ* expression becomes undetectable in the mutant clones (arrows) is indistinguishable from that in wild-type tissue (arrowheads). *brk* clones are revealed by the absence of GFP (**n**; green). **o**, Superimposed images of **m** and **n**. The amount of *brk-lacZ* expression in *brk* clones is higher than in wild-type tissue because two copies of *brk-lacZ* are present in cells of the *brk* clone, whereas only one copy of *brk-lacZ* is present outside the clone, and no *brk-lacZ* is present in the twin spot (**m**: enclosed by dotted line). **p**, Model relating Brk activity and Dpp signalling. *brk* is negatively regulated by Dpp signalling that is transduced by *tkv* and *Mad*, suggesting that *brk* does not function upstream of *Mad*, but may function parallel to or downstream of *Mad* when regulating Dpp-target genes. Arrows denote upregulation; tee bars, downregulation.

globin mRNA¹⁷. We found that co-injection of *brk* and *Mad* mRNA nullified the effects caused by injection of each mRNA alone (Fig. 4d), indicating that Brk can negatively regulate Dpp/BMP-4 signalling by competing with *Mad* and that a vertebrate homologue of Brk may exist.

Our results demonstrate that Brk represses the transcriptional targets of Dpp in the wing disc and that Dpp activates transcriptional targets by repressing *brk*, a model that is consistent with the presumed distribution of Dpp and the observed expression pattern of *brk*. Dpp is thought to be present as a gradient, with its highest

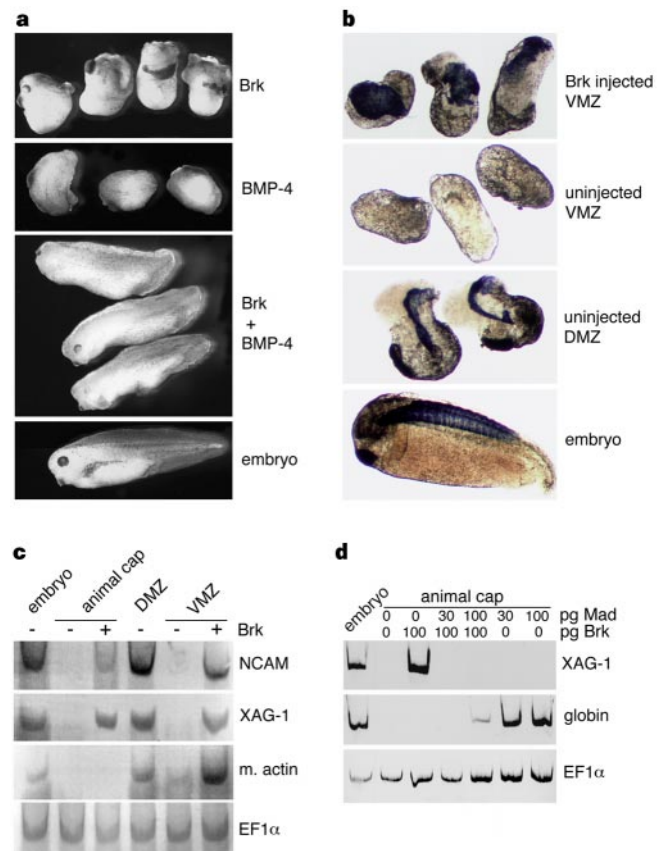


Figure 4 Brk can antagonize signalling by BMP-4. **a**, Ectopic expression of *brk* mRNA dorsalizes *Xenopus* embryos and suppresses ventralization by BMP-4. Two-cell embryos were injected with 200 pg synthetic *brk* RNA (Brk), 200 pg synthetic BMP-4 RNA (BMP-4) or 200 pg each of *brk* and BMP-4 RNAs (Brk + BMP-4) into both blastomeres and then cultured to the tadpole stage. **b**, Brk induces muscle formation in ventral marginal zone (VMZ) explants. VMZ and dorsal marginal zone (DMZ) explants were cultured until sibling embryos reached stage 30 and were immunostained as described²² with the muscle-specific antibody 12/101 (ref. 23). **c**, Brk induces neural tissue and dorsalizes ventral mesoderm. For animal cap assays, 200 pg *brk* mRNA was injected near the animal pole of two-cell embryos and animal caps were explanted at the blastula stage. For marginal zone explants, 200 pg *brk* mRNA was injected into two ventral blastomeres of four-cell embryos. DMZ and VMZ explants were isolated at stage 10.5 (staging of the embryo was done according to ref. 24). Explants were cultured until sibling embryos reached stage 26 and were analysed for expression of mesodermal (m. actin), neural (NCAM) and cement gland (XAG-1) specific genes by reverse transcription with polymerase chain reaction (RT-PCR) as described²⁵, except that PCR products were detected by silver staining of a 5% polyacrylamide gel. The sequences of EF1-α, NCAM and muscle actin²⁶ and XAG-1 (ref. 27) have been reported previously. **d**, Brk antagonizes *Mad* activity. For animal cap assays, embryos were injected with synthetic RNA encoding *Mad* or Brk, as indicated. After injection, embryos were treated as in **c** and analysed for expression of cement gland (XAG-1) and ventral mesodermal (globin) specific genes by RT-PCR. Co-injection of *Mad* (30 pg) and *brk* (100 pg) synthetic RNAs cancels the effects caused by injection of each mRNA alone.

concentration being at the A/P border and gradually declining towards the periphery. We believe that this distribution of Dpp generates a complementary gradient of Brk-repressing activity. □

Methods

Cloning of the *brk* gene. Genomic DNA adjacent to the P-element transposon in strain no. 38 (from a collection of Y.-N. Jan) was isolated by plasmid rescue. A 3.7-kb fragment of the rescued plasmid was used to isolate genomic clones from a λ -phage library (a gift from Y.-N. Jan). Genomic fragments containing exons were identified by northern blotting and fragments that hybridized to RNA species with a distribution identical to the expression pattern of β -galactosidase in strain 38 were used to screen a 4–8-h embryonic cDNA library (from N. Brown)¹⁸. The longest cDNA clone, e7-4, was sequenced on both strands using an automated sequencer.

Ectopic expression. Transgenes and the enhancer trap line used were *UAS-dpp* and *vg-Gal4* (a gift from S. Morimura), *omb^{P1}* (ref. 5), *sal^{A405.1M2}* (ref. 19), *30A-Gal4* (ref. 11), *3sB-Gal4* (F. A. Ramirez-Weber and T. Kornberg, unpublished results). Discs from the larvae of the genotype *30A-Gal4/+;UAS-lacZ/+* (Fig. 1h) and *#38/+;30A/+;UAS-dpp/+* (Fig. 1i) were stained for β -galactosidase activity. *dpp*-expressing clones (Fig. 3c) have been described².

Clonal analysis. Clones of cells mutant for *Mad* were made by using the Flp-FRT technique²; for *brk*, they were made either by the Flp-FRT technique for immunostaining, or by γ -ray irradiation (1,350 rads from a ¹³⁷Cs source) for analysing wing phenotypes. Discs from larvae of the genotype *hsflp/brk-lacZ;FRT40Mad¹⁻²/FRT40Myc* (Fig. 1k, l), *brk³⁸⁻³FRT18A/Ub-GFP FRT18A;flp388/+* (Fig. 3d–f), *brk³⁸⁻³FRT18A/Ub-GFP FRT18A;sal^{A405.1M2}/+;flp388/+* (Fig. 3d–f), *brk³⁸⁻³FRT18A/Ub-GFP FRT18A;flp388/P1883* (Fig. 3j–l) and *brk³⁸⁻²⁰FRT18A/Ub-GFP FRT18A;flp388/+* were immunostained for β -galactosidase (Figs 1i, j, 3g–l) or Omb (Fig. 3j–l), and then observed on a Zeiss confocal laser microscope 510. The *Ub-GFP FRT18A* chromosome was made by recombination between *Ub-GFP FRT10.1* (ref. 20) and *FRT18A* (ref. 21).

Received 4 January; accepted 5 February 1999.

1. Capdevila, J. & Cuerrero, I. Targeted expression of the signaling molecule decapentaplegic induces pattern duplications and growth alterations in *Drosophila* wings. *EMBO J.* **13**, 4459–4468 (1994).
2. Zecca, M., Basler, K. & Struhl, G. Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the *Drosophila* wing. *Development* **121**, 2265–2278 (1995).
3. Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a DPP morphogen gradient. *Cell* **85**, 357–368 (1996).
4. Lecuit, T. *et al.* Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. *Nature* **381**, 387–393 (1996).
5. Grimm, S. & Pflugfelder, G. O. Control of the gene optomotor-blind in *Drosophila* wing development by decapentaplegic and wingless. *Science* **271**, 1601–1604 (1996).
6. de Celis, J. F., Barrio, R. & Kafatos, F. C. A gene complex acting downstream of *dpp* in *Drosophila* wing morphogenesis. *Nature* **381**, 421–424 (1996).
7. Tsuneizumi, K. *et al.* Daughters against *dpp* modulates *dpp* organizing activity in *Drosophila* wing development. *Nature* **389**, 627–631 (1997).
8. Inoue, H. *et al.* Interplay of signal mediators of Decapentaplegic (Dpp): Molecular characterization of Mothers against *dpp*, Medea, and Daughters against *dpp*. *Mol. Biol. Cell* **9**, 2145–2156 (1998).
9. Jazwinska, A., Rushlow, C. & Roth, S. *brk*, a component of the *dpp* pathway, affects patterning of the *Drosophila* appendages. *A. Conf. Dros. Res.* **39**, 244A (1998).
10. Rushlow, C., Silver, S. & Roth, S. *brinker*, a negative regulator of the *dpp* pathway. *A. Conf. Dros. Res.* **39**, 245B (1998).
11. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
12. Sekelsky, J., Newfeld, S., Raftery, L., Chartoff, E. & Gelbart, W. Genetic characterization and cloning of Mothers against *dpp*, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics* **139**, 1347–1358 (1995).
13. Burtis, K. C. & Baker, B. S. *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56**, 997–1010 (1989).
14. Ferguson, E. L. Conservation of dorsal–ventral patterning in arthropods and chordates. *Curr. Opin. Genet. Dev.* **6**, 424–431 (1996).
15. Kao, K. R. & Elinson, R. P. Dorsalization of mesoderm induction by lithium. *Dev. Biol.* **132**, 81–90 (1989).
16. Graff, J. M. Embryonic patterning: To BMP or not to BMP, that is the question. *Cell* **89**, 171–174 (1997).
17. Newfeld, S. J., Chartoff, E. H., Graff, J. M., Melton, D. A. & Gelbart, W. M. Mothers against *dpp* encodes a conserved cytoplasmic protein required in DPP/TGF- β responsive cells. *Development* **122**, 2099–2108 (1996).
18. Brown, N. H. & Kafatos, F. C. Functional cDNA libraries from *Drosophila* embryos. *J. Mol. Biol.* **203**, 425–437 (1988).
19. Wagner-Bernholz, J. T., Wilson, C., Gibson, G., Schuh, R. & Gehring, W. J. Identification of target genes of the homeotic gene *Antennapedia* by enhancer detection. *Genes Dev.* **5**, 2467–2480 (1991).
20. Duffy, J. B., Harrison, D. A. & Perrimon, N. Identifying loci required for follicular patterning using directed mosaics. *Development* **125**, 2263–2271 (1998).
21. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
22. Hemmati-Brivanlou, A. & Harland, R. M. Expression of an engrailed-related protein is induced in the anterior neural ectoderm of early *Xenopus* embryos. *Development* **106**, 611–617 (1989).

23. Kintner, C. & Brookes, J. P. Monoclonal antibodies identify blastemal cells derived from differentiating muscle in new limb regeneration. *Nature* **308**, 67–69 (1984).
24. Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis* (Garland, New York and London, 1994).
25. Wilson, P. A. & Melton, D. A. Mesodermal patterning by an inducer gradient depends on secondary cell–cell communication. *Curr. Biol.* **4**, 676–686 (1994).
26. Hemmati-Brivanlou, A. & Melton, D. A. Inhibition of activin receptor signaling promotes neuralization in *Xenopus*. *Cell* **77**, 273–281 (1994).
27. Blitz, I. L. & Cho, K. W. Y. Anterior neurectoderm is progressively induced during gastrulation: the role of the *Xenopus* homeobox gene *orthodenticle*. *Development* **121**, 993–1004 (1995).

Acknowledgements. We thank C. Rushlow for exchanging unpublished data; T. Kornberg and J. Christian for critically reading the manuscript; H. Eguchi in Research Center for Nuclear Science and Technology for help with γ irradiation experiments; K. Niwano for technical assistance; members of T.T.'s laboratory for their help; K. Basler, S. Goto, Y. N. Jan, T. Kornberg, M. Mlodzik, S. Morimura, G. Pflugfelder, F.-A. Ramirez-Weber, F. Roth, S. Roth, G. Struhl and the Bloomington *Drosophila* Stock Center for fly strains; G. Pflugfelder for the Omb antibody; and M. Kirschner and S. Newfeld for BMP-4 and Mad cDNA, respectively. This work was supported by grants from the Japan Society for the Promotion of Science (Research for the Future Program) to T.T. and grants-in-aid from the Ministry of Education, Science and Culture of Japan to M.M., N.K. and T.T. M.M. is a Research Fellow of the Japan Society for the Promotion of Science.

Correspondence and requests for materials should be addressed to T.T. (e-mail: ttataba@ims.u-tokyo.ac.jp). DDBJ sequence accession number is AB023583

A new protease required for cell-cycle progression in yeast

Shyr-Jiann Li & Mark Hochstrasser

University of Chicago, Department of Biochemistry & Molecular Biology, 920 East 58th Street, Chicago, Illinois 60637, USA

In eukaryotes, protein function can be modulated by ligation to ubiquitin or to ubiquitin-like proteins (Ubl proteins)^{1–3}. The vertebrate Ubl protein SUMO-1 is only 18% identical to ubiquitin but is 48% identical to the yeast protein Smt3. Both SUMO-1 and Smt3 are ligated to cellular proteins, and protein conjugation to SUMO-1/Smt3 is involved in many physiological processes^{3–10}. It remained unknown, however, whether deconjugation of SUMO-1/Smt3 from proteins is also essential. Here we describe a yeast Ubl-specific protease, Ulp1, which cleaves proteins from Smt3 and SUMO-1 but not from ubiquitin. Ulp1 is unrelated to any known deubiquitinating enzyme but shows distant similarity to certain viral proteases, indicating the existence of a widely conserved protease fold. Proteins related to Ulp1 are present in many organisms, including several human pathogens. The pattern of Smt3-coupled proteins in yeast changes markedly throughout the cell cycle, and specific conjugates accumulate in *ulp1* mutants. Ulp1 has several functions, including an essential role in the G2/M phase of the cell cycle.

The ubiquitin and Ubl protein modifiers are linked by an amide (isopeptide) bond between the carboxy terminus of the modifier and a lysine side chain(s) of the substrate. Both modifiers are also synthesized in precursor form: one or more amino acids follow a glycine–glycine dipeptide that will form the C terminus of the mature protein. For ubiquitin precursors, these tail sequences are clipped off by deubiquitinating (Dub) enzymes¹⁰. The Dub enzymes are a heterogeneous group of specialized cysteine proteases that are also capable of precisely cleaving ubiquitin from ubiquitin–protein fusions and/or isopeptide-bond-linked ubiquitin–protein conjugates¹⁰. No Ubl-specific processing enzyme had, to our knowledge, been identified until now.

To isolate a *Saccharomyces cerevisiae* Smt3-cleaving enzyme, we first generated a series of Smt3-based substrates for expression in *Escherichia coli*. We then biochemically screened for Smt3-cleaving enzymes among yeast proteins expressed in *E. coli* from a genomic DNA library¹¹. Using a sib-selection protocol¹², we assayed cleavage of a ³⁵S-labelled substrate consisting of six histidine residues, ubiquitin, Smt3 and haemagglutinin (HA) (Fig. 1a) in extracts made from pools of bacterial transformants. We analysed reaction products by SDS–polyacrylamide gel electrophoresis, which allowed us to follow cleavage after the glycine–glycine sequences

in both ubiquitin and Smt3. From an initial set of pools made from $\sim 1.6 \times 10^4$ colonies, we identified three pools with Smt3-cleaving activity and 22 pools with ubiquitin-cleaving activity. (Only one of the pools was capable of cleavage after both ubiquitin and Smt3; ubiquitin processing activity from this pool was traced to a library plasmid bearing the *UBP1* gene¹².) Further screening of the three initial Smt3-cleavage-positive pools identified the transformants responsible for Smt3 processing. Plasmids isolated from these transformants contained overlapping DNA fragments from yeast chromosome 16. We used DNA subcloning to trace the Smt3-cleaving activity to the *YPL020c* open reading frame (ORF). *YPL020c* is a previously uncharacterized ORF; the predicted protein, of relative molecular mass 72,000 (M_r 72K), shows no evident similarity to any Dub (Fig. 2). We have designated the product of this gene Ulp1 (for Ubl-specific protease 1).

We investigated the enzymatic properties of Ulp1 by using a glutathione-S-transferase (GST)-Ulp1 fusion protein purified

from *E. coli*. To confirm that Ulp1 cleaves specifically at the C terminus of Smt3, we incubated purified Smt3- β -galactosidase with the enzyme, and subjected the β -galactosidase product to amino-terminal peptide sequencing. The sequence matched the N-terminal sequence predicted after precise removal of Smt3 (data not shown). Ulp1 also efficiently processed the natural Smt3 precursor (Fig. 1d, lanes 3–5). Cleavage by Ulp1 was completely blocked by preincubation of the enzyme with 5 mM *N*-ethylmaleimide, a thiol reagent, but was unaffected by high concentrations of the Dub inhibitor ubiquitin aldehyde (1 μ M) or the serine protease inhibitor phenylmethylsulphonyl fluoride (1 mM) (Fig. 1b). These results are consistent with Ulp1 being a cysteine protease that is distinct from a Dub enzyme. Kinetic analysis of Ulp1 using Smt3-HA and Smt3- β -galactosidase showed that both substrates were cleaved efficiently (Table 1), indicating that Ulp1 is capable of processing diverse Smt3-protein substrates.

We also tested the activity of Ulp1 against an isopeptide-linked substrate by measuring cleavage of the mammalian SUMO-1 protein (also called UBL1, PIC1, sentrin, SMT3C or GMP1) from Lys 526 of ranGAP1 (ref. 13), a GTPase-activating protein that regulates Ran activity. ranGAP1 is required for nucleocytoplasmic transport, and its modification by SUMO-1 is required for ranGAP1 localization to the nuclear pore. Radiolabelled mouse ranGAP1 was synthesized in a rabbit reticulocyte-lysate-derived transcription/translation system; $\sim 30\%$ of the ranGAP1 protein was conjugated to SUMO-1 present in the lysate (Fig. 1c, lane 2). A mutant ranGAP1, in which the lysine that participates in the isopeptide bond to SUMO-1 was changed to arginine, could not be conjugated to SUMO-1, as expected (ref. 13 and Fig. 1c, lane 1). To test for cleavage activity, we incubated translation products with GST-Ulp1; this resulted in quantitative cleavage of the ranGAP1-SUMO1 conjugate (Fig. 1c). Finally, endogenous yeast Smt3-protein conjugates present in a crude extract could also be efficiently cleaved by GST-Ulp1 (Fig. 1d). These results show first, that yeast Ulp1 can process both Smt3- and SUMO1-linked proteins, and second, that Ulp1 can cleave both peptide and isopeptide bonds between Smt3/SUMO-1 and substrate.

Database searches identified a number of sequences related to Ulp1 (Fig. 2), including another *S. cerevisiae* protein, Smt4 (ref. 14). The *SMT4* gene was identified in the same high-copy-suppressor screen in which *SMT3* was detected¹⁴, although no functional studies of *SMT4* have been published. The similarity between the 621-residue Ulp1 and 1,034-residue Smt4 proteins is confined primarily to a region of ~ 200 amino acids, and this region is also the part most readily aligned with other Ulp1-related proteins from other organisms (Fig. 2a). On the basis of the inhibitor studies, we searched for conserved residues in this domain that could be components of a cysteine-protease active site. We identified one conserved cysteine and one conserved histidine residue in the aligned proteins (arrowheads in Fig. 2a). When we used the domain containing these invariant residues (Fig. 2a) in a PSI-BLAST¹⁵ sequence search, we detected weak similarity to gene products from several animal viruses, including the processing proteases of many adenoviruses (Fig. 2b). A 2.6 Å crystal structure is available for the human type-2 adenovirus protease¹⁶. In the structure, the residues implicated in catalysis can be superimposed on the corresponding active-site residues of the classical cysteine protease papain, despite the absence of primary-sequence similarity

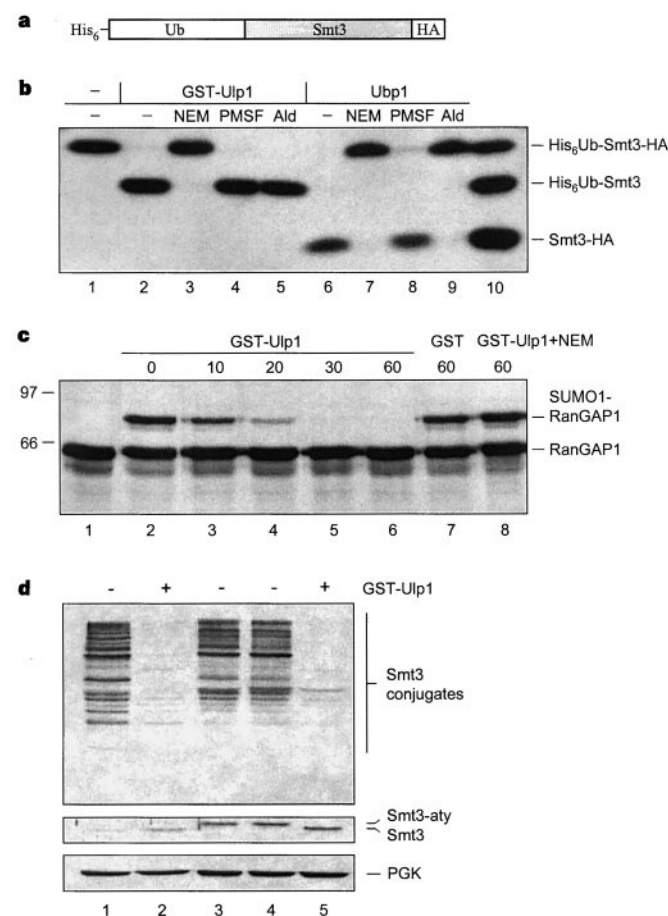


Figure 1 Characterization of the Ulp1 enzyme. **a**, The His₆-ubiquitin-Smt3-HA substrate. Ub, ubiquitin. **b**, Cleavage of His₆-Ub-Smt3-HA by Ulp1 and Ubp1, analysed by SDS-PAGE. Lane 1, no added enzyme; lane 10, mixture of separately purified proteins. Preincubations with inhibitors (NEM, PMSF and Ald, for Ub aldehyde) were done for 15 min at room temperature. Ubp1 was used as a control for cleavage after Ub in the substrate and to confirm the inhibitory activity of Ub aldehyde. **c**, Cleavage of SUMO1-ranGAP1 synthesized *in vitro*. Lane 1, ranGAP1-K526R mutant; lanes 2–6, human ranGAP1 incubated at 30 °C with purified GST-Ulp1 for indicated times (min); lane 8, GST-Ulp1 pretreated with NEM. Relative molecular mass standards are indicated (in thousands). **d**, *In vitro* cleavage by Ulp1 of yeast Smt3-protein conjugates. A longer exposure was needed to visualize free Smt3. Yeast extracts (25 μ g) were from *ulp1-ts* cells grown at 23 °C (lanes 1, 2) and at 37 °C (lanes 3–5); lane 3, GST added. Immunoblot used a rabbit anti-Smt3 antibody raised against His₆-Smt3. PGK, phosphoglycerate kinase; Smt3-aty, Smt3 precursor; Smt3, mature Smt3.

Table 1 Cleavage of Smt3-fusion proteins by GST-Ulp1

Substrate	V_{obs} (min ⁻¹)
His ₆ -Smt3-HA	45
His ₆ -Smt3-Met- β -galactosidase	17
His ₆ -Smt3-Leu- β -galactosidase	11

Substrate turnover numbers (V_{obs}) were determined under initial rate conditions at 30 °C. The concentration of GST-Ulp1 was 0.45 nM and substrates were present in 10^3 – 10^4 molar excess over enzyme. Duplicate measurements at each reaction time point were made.

between the enzymes. All four key catalytic residues of the adenovirus protease are fully conserved in the Ulp1-related enzymes (arrowheads in Fig. 2a, b), indicating that the Ulp proteins may use a catalytic mechanism similar to that of the viral protease. Adenovirus, African swine fever virus and the pox viruses all process viral-protein precursors at consensus sites that match or are closely related to the glycine-glycine-X cleavage site of Smt3 and SUMO-1 precursors^{17,18}.

To test the predicted catalytic similarities between adenovirus protease and Ulp proteins, we studied Smt3 processing by purified wild-type and mutant GST-Ulp1 fusions; in the mutants, the conserved cysteine and histidine residues were mutated to serine and arginine residues, respectively (Fig. 3a). No cleavage was detected in the presence of ulp1-S580 after 3 h (Fig. 3a, lane 4), and extremely weak activity was detected for ulp1-R514 (lanes 1, 2). Therefore, the cysteine and histidine predicted by sequence alignment to be active-site residues are indeed important for catalytic activity. We propose that the ~90-residue segment shown in Fig. 2b forms a core structure common to a diverse and widespread group of cysteine proteases. Except for the adenovirus protease, the proteins shown in Fig. 2b have not previously been shown to be proteases. Extra regions of similarity among the proteins in Fig. 2a,

especially evident upstream of the core domain, may contribute to selectivity for Smt3/SUMO-1.

To investigate the physiological functions of *ULP1*, we deleted one of the two copies of the gene in a diploid yeast strain. Sporulation and tetrad analysis of the resulting heterozygotes (20 full tetrads) showed that the *ulp1Δ* segregants were inviable. Microscopic examination of the *ulp1Δ* segregants over several days showed that growth arrested asynchronously, resulting in microcolonies of ~50–100 cells. Interestingly, 85–90% of the cells arrested as large-budded cells (which eventually became enlarged and often lysed), indicating that cell-cycle-specific arrest may have occurred (see later). To confirm that *ULP1* is required for vegetative growth and not just development from spores, we transformed the *ulp1Δ* heterozygote with YCp50-ULP1, a low-copy plasmid carrying both *ULP1* and *URA3*; the transformants were sporulated and dissected. Viable *ulp1Δ* segregants containing YCp50-ULP1 were identified. When these cells were grown on 5-fluoro-orotic acid (FOA), which is toxic to cells expressing *URA3*, no colonies were isolated at either 23 °C or 30 °C, indicating that *ULP1* is indeed essential for vegetative growth (Fig. 3b). Growth on FOA could be restored if wild-type *ULP1* were present on a second plasmid that did not contain *URA3*. In contrast, the catalytically inactive *ulp1-*



Figure 2 Ulp1 is a new type of cysteine protease. **a**, Sequence alignment of a highly conserved region of ~180–220 residues in Ulp1 and representative Ulp1-related proteins. Arrowheads mark likely residues of the catalytic triad (histidine, aspartate and cysteine) and an invariant glutamine residue predicted to help form the oxyanion hole in the active site. The *S. cerevisiae* (Sc) Ulp1 and Smt4 proteins are shown, as are segments from predicted proteins from humans (HsUlp1), mouse (MmUlp1), *S. pombe* (SpUlp1, SpUlp2), *Drosophila* (DmUlp1), rice

(OsSEB1) and *Plasmodium falciparum* (PfUlp1). Similar proteins from many other organisms are also easily detected in the databases. **b**, Sequence alignment of the 'core domains' of yeast Ulp1, a predicted protein (CtMTP) from *Chlamydia* (GenBank accession number AE001360), the adenovirus L3 protease (Ad2can; accession number U77082), and ASFV ORF (accession number X71982), and I7 proteins of fowlpox (accession number F48563) and vaccinia virus (accession number P12926).

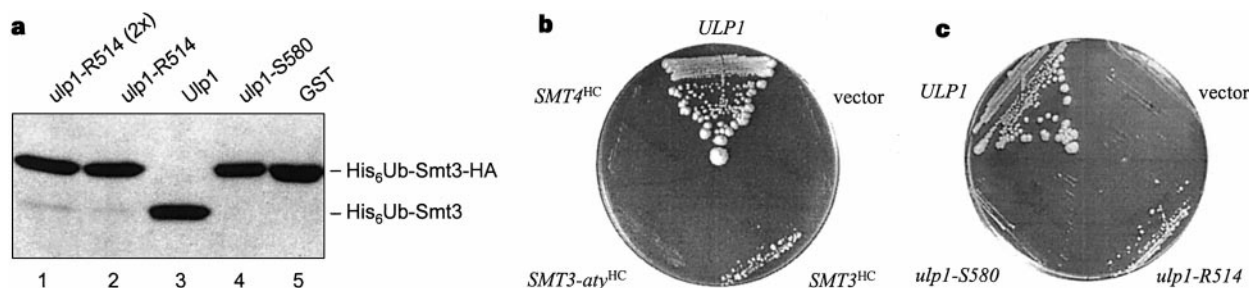


Figure 3 Mutagenesis of putative Ulp1 active-site residues and characterization of *ulp1* mutants. **a**, Cleavage activity of the *ulp1*-S580 and *ulp1*-R514 mutants. Purified GST derivatives were incubated with radiolabelled substrate at 30°C. Double the amount of the reaction with GST-*ulp1*-R514 was loaded in the first lane

to allow visualization of product. **b**, Growth on FOA of *ulp1*Δ[YCp50-ULP1] cells transformed with high-copy-number (HC) plasmids bearing the indicated genes. The plate was incubated at 30°C for 15 days. **c**, As for **b** except that the incubation time was 11 days.

S580 allele failed to rescue growth, and the *ulp1*-R514 allele, which encodes a protein with extremely low residual Smt3-cleaving activity (Fig. 3a), allowed very slow growth on FOA (Fig. 3c). Thus, the Smt3-cleaving activity of Ulp1 correlates closely with its ability to support yeast growth.

If Ulp1 were the only, or at least the main, Smt3-precursor-processing enzyme in yeast, the inviability of the *ulp1*Δ strain could be explained by a failure to process sufficient Smt3 precursor. To test this possibility, we transformed *ulp1*Δ [YCp50-ULP1] cells with plasmids encoding either the normal Smt3 precursor (Smt3-*aty*) or the mature form of the protein (Smt3). Following eviction of the *ULP1*-bearing plasmid, no growth of cells expressing Smt3-*aty* was detectable, whereas very small colonies of cells overproducing the

mature Smt3 protein could be seen after 1–3 weeks at 23 °C or 30 °C (Fig. 3b). (Both plasmids fully rescued growth of an *smt3*Δ strain (data not shown).) These results indicate that Ulp1 contributes to Smt3-precursor processing *in vivo*, but because cells supplied with mature Smt3 were still severely growth-impaired, Ulp1 also must be responsible for cleaving Smt3 from post-translational conjugates (see later). This latter function is critical for growth.

To determine whether a cell-cycle-specific growth arrest was seen in vegetative cultures depleted of Ulp1, we first studied *ulp1*Δ cells that expressed *ULP1* from a plasmid under the control of a galactose-inducible promoter. When transcription of *GAL-ULP1* was shut off by switching cells to glucose medium for 12 h, cell division also ceased, and cells accumulated in large-budded form, in

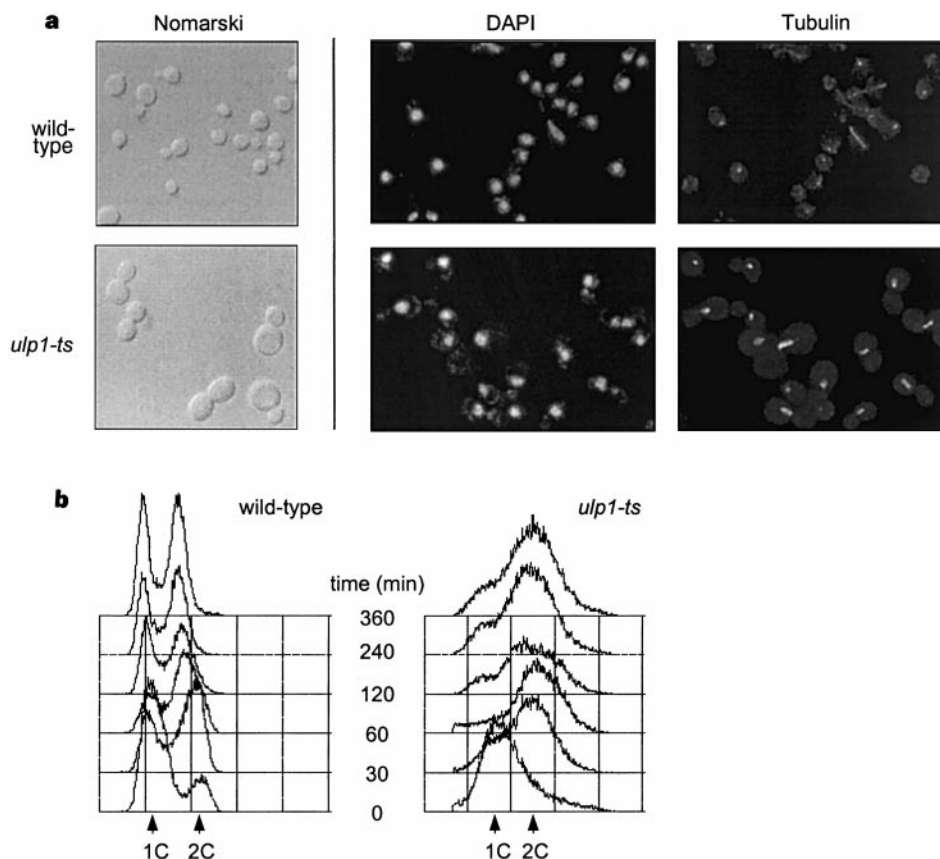


Figure 4 Ulp1 is required for cell-cycle progression. **a**, Cell-cycle phenotype of *ulp1*-*ts* cells. Micrographs show cell morphology (Nomarski) after 6 h at 37°C and, for a different set of samples, nuclear staining (with DAPI) and microtubule immunostaining after 8.5 h at 37°C. Mutant cells are enlarged compared with wild-type cells. After 8 h in YPD medium at 37 °C, 58.5 ± 3.5% of mutant cells (*n* = 572)

had large buds, 7.2 ± 1.8% had small buds and 34.3 ± 4.1% had no buds. The corresponding percentages for wild-type cells (*n* = 425) were 32 ± 1.8, 4 ± 1.7 and 46.6 ± 1.9. **b**, FACS analysis of cells synchronized with α -factor and released at 37°C. 1C and 2C refers to one and two sets of chromosomes per cell.

contrast to cells left in galactose medium (data not shown). For a more detailed analysis of the apparent cell-cycle defect, we isolated temperature-sensitive (*ts*) *ulp1* alleles. Cells bearing one of these alleles had a slight growth defect at 23 °C and could not form colonies at 37 °C. Cell proliferation in liquid culture was severely reduced within a few hours at 37 °C. We synchronized wild-type and *ulp1-ts* cells grown at 23 °C in G1 phase by incubating them with α -factor. After removal of the pheromone, the cultures were shifted to 37 °C and cells were analysed by fluorescence-activated cell sorting (FACS) and light microscopy (Fig. 4a, b). Before the shift, both mutant and wild-type cells were predominantly in G1 phase, as judged by their haploid DNA content and unbudded cell morphology. Both mutant and wild-type cells entered S phase and G2/M phase with comparable kinetics. However, most *ulp1-ts* cells remained in G2/M, whereas wild-type cells went on to divide and re-enter G1 within a few hours. Staining of nuclei with 4,6-diamidino-2-phenylindole (DAPI) and an anti-tubulin antibody revealed a single nucleus near the bud neck and a short spindle in most of the mutant cells (Fig. 4a), indicating that the cells had accumulated at a stage before anaphase. After prolonged incubation at restrictive temperature, distinct subdomains of nuclear staining with propidium iodide were often seen, indicating either chromosome fragmentation or other alterations in chromosome structure; this would be consistent with the reproducibly broader distribution of fluorescence intensities seen in FACS analysis of the mutant (Fig. 4b). Overproduction of mature Smt3 in *ulp1-ts* cells did not alter the FACS profile detectably (data not shown), indicating that the cell-cycle arrest was not due to a limiting amount of Smt3 that could be conjugated to substrates.

By anti-Smt3 immunoblot analysis, several Smt3-containing species could be discerned in exponentially growing cultures and in cultures approaching stationary phase; the two patterns, however, were distinct (Fig. 5b). Very little free Smt3 was detected under any condition. When wild-type cells were arrested at different stages of the cell cycle or were followed after synchronization with α -factor-mediated arrest, the Smt3-conjugate pattern underwent marked changes as a function of cell-cycle position (Fig. 5b). To verify that the high-molecular-mass species that reacted with the anti-Smt3 antibody were in fact Smt3-protein conjugates, we studied a *ubc9-ts* mutant, which is defective for an enzyme that conjugates Smt3 to proteins¹⁹. In *ubc9-ts* cells, particularly at 37 °C, a large decrease in Smt3 conjugates and an increase in free Smt3 were observed (Fig. 5c). In *ulp1-ts* cells at the permissive temperature, a marked accumulation of specific Smt3 conjugates was seen relative to wild-type cells (Fig. 5c). Conjugates also accumulated in *ulp1Δ* cells that overproduced mature Smt3 (Fig. 5a). Paradoxically, at 37 °C the levels of detectable Smt3 conjugates decreased when *ulp1-ts* cells were lysed directly in SDS loading buffer (Fig. 5c). However, this decrease was not seen when cells were converted to spheroplasts and lysed by sonication (Fig. 1d, compare lanes 1, 3, 4). These results suggest that the extractability of Smt3-protein conjugates changes in *ulp1-ts* cells when they are shifted to high temperature. We conclude that the pattern of Smt3-protein modification *in vivo* is dynamic and regulated by cell-cycle and growth stage. Moreover, the pattern of conjugates depends on Ulp1.

The ability of large amounts of mature Smt3 to rescue (weakly) the lethality of the *ulp1Δ* mutant indicates a low-frequency bypass of the G2/M-phase cell-cycle block. Under these conditions, mobilization of Smt3 from conjugates is not required to supply free Smt3 for further ligation reactions. The extremely poor growth of these cells indicates that detachment of Smt3 from one or more proteins by Ulp1 is required for normal cell-cycle progression and/or that accumulation of abnormal Smt3 conjugates inhibits growth. However, conjugation of Smt3 to substrates must also positively regulate the G2/M transition, as *ubc9* mutants show a similar cell-cycle arrest⁵. Different targets of Smt3 conjugation may account for these negative and positive cell-cycle-regulatory effects, but it is also

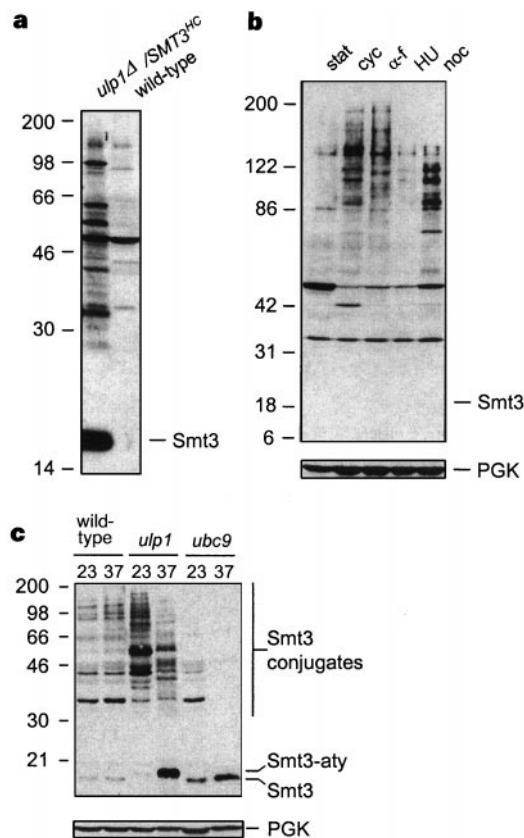


Figure 5 Smt3-protein conjugates in yeast. **a**, Smt3-conjugate accumulation in *ulp1Δ* cells overproducing His₆-Smt3. **b**, Smt3 conjugates in cells grown under different conditions. Stat, stationary phase; cyc, cycling cells; α -f, plus α -factor; HU, plus hydroxyurea; noc, plus nocodazole. Arrest points were verified by FACS analysis and microscopy. **c**, Anti-Smt3 immunoblot analysis of logarithmic cultures grown at 23 °C and then either shifted to 37 °C or left at 23 °C for an extra 6 h. (The region showing free Smt3 is from a longer exposure of the same gel.)

possible that the attachment of Smt3 to a particular protein(s) is needed early in G2/M, and that removal of Smt3 from the same protein(s) is required subsequently. In either case, it seems likely that Ulp1 activity will be closely regulated to ensure the proper dynamics of these modifications.

Although Smt3-protein ligation is essential for yeast viability and cell-cycle progression^{3,5,6}, it does not necessarily follow that Smt3-protein deconjugation must also be critical. In an earlier study, a cleavage-resistant Smt3 mutant supported wild-type rates of yeast growth, leading the authors to conclude that Smt3-protein deconjugation is not essential in yeast⁶. Moreover, although conjugation of substrates to ubiquitin is critical for cell-cycle progression, none of the 17 Dub-encoding genes in yeast is essential^{20,21}. Our demonstration that Ulp1-mediated Smt3-protein deconjugation is indeed required for cell-cycle progression recalls the requirement for timed phosphorylation and dephosphorylation of cell-cycle proteins²². Cell-cycle progression may involve a dynamic interplay between these two types of post-translational modification. The accumulation of specific Smt3-protein conjugates in *ulp1* strains should facilitate identification of the targeted substrates; this will be needed to determine the exact mechanistic roles of Smt3.

Manipulation of Ulp activities is also likely to prove useful in mammalian cells. SUMO-1 can negatively regulate the NF- κ B pathway by being conjugated to I κ B, thus blocking ubiquitination of I κ B⁹. Because the NF- κ B pathway is central to inflammatory and immune responses, alteration of Ulp enzymatic activity could be

used to modify these responses when they are inappropriately activated. There may be a connection between a defect in SUMO-1 conjugation to the PML protein and acute promyelocytic leukaemia (APL)²³. In APL cells, PML is fused to retinoic-acid receptor- α (RAR), and the PML-RAR fusion is no longer modified by SUMO-1. Modification of PML-RAR is observed upon overproduction of SUMO-1 (ref. 23). Treatment of APL cells with arsenic trioxide also increases modification of PML by SUMO-1 and is an effective treatment for APL patients. Inhibition of the enzyme responsible for SUMO-1-PML deconjugation has been suggested as an explanation for these effects²⁴. More specific Ulp inhibitors could therefore have therapeutic value for APL patients. □

Methods

Activity screen for Smt3-processing enzymes. *E. coli* strain HB101 transformed with a yeast genomic library¹¹ was plated at a density of ~50 colonies per plate on ampicillin medium (LBamp). Cells were replica-plated onto the same medium and then collected from the replica plates with 2 ml LBamp. Following 2 h of growth at 37 °C, cells were pelleted, resuspended in 50 μ l spheroplasting solution (25% sucrose, 0.25M Tris-HCl, pH 8.0, 2 mg ml⁻¹ lysozyme), and incubated for 5 min on ice. 150 μ l lysis buffer (330 mM Tris-HCl, pH 8.0, 75 mM EDTA, 1 mM PMSE, 2 mM dithiothreitol (DTT)) was added, and the incubation was continued on ice for an extra 5 min. After a brief sonication, extracts were centrifuged at 14,000g for 15 min and 15 μ l of the supernatant was used for incubation for 3 h at 37 °C with ³⁵S-labelled His₆-ubiquitin-Smt3-HA affinity-purified substrate. When a pool with Smt3-cleaving activity was identified, the colonies from the master plate were subdivided and then treated as above to identify the subpools with Smt3-cleaving activity.

Plasmid and strain constructions. Standard methods were used for growth and manipulation of yeast and bacteria²⁵. To generate yeast strains with null alleles of *SMT3* or *ULP1*, the yeast *HIS3* gene was amplified by the polymerase chain reaction (PCR) using primers with 5'-sequence segments matching the beginning and end of the respective ORF sequences; the amplified fragments were transformed into diploid MHY606 cells²⁰, and the resulting histidine prototrophs were checked by colony PCR for the correct gene replacements.

Several plasmids for expression of Smt3 derivatives from the *CUP1* promoter in yeast were constructed. PCR products encoding His₆-tagged precursor or mature Smt3 protein were ligated into plasmid YRTAG310 (GenBank accession number U37458). Low-copy plasmids encoding the N-terminally HA-tagged derivatives of precursor and mature Smt3 were made by cloning the corresponding PCR fragments into plasmid YATAG200 (GenBank accession number U37457). The *PvuII* fragments from these two plasmids were also subcloned into *SmaI*-cut pRS424. The *E. coli* expression vector pQE30 (Qiagen) was used for expression and purification of His₆-tagged proteins. Details of plasmid constructions are available upon request.

Site-directed and random mutagenesis of ULP1. Mutations in codons encoding putative active-site residues of Ulp1 were introduced into a YCplac22-*ULP1* template using the QuikChange system (Stratagene). *Kpn21-SacI* fragments containing the mutagenized codons were sequenced in their entirety and were used to replace the corresponding wild-type fragment in YCplac22-*ULP1* and pRS424-*ULP1* for expression in yeast and in pGEX-*ULP1* for expression as a GST-fusion protein in *E. coli*. A library of mutant alleles of *ULP1* was made by PCR-mediated mutagenesis²⁶. Plasmid DNA was isolated from the library and used to transform yeast strain MHY1322 (*ulp1 Δ :HIS3*[YCp50-*ULP1*]). Colonies were replica-plated in duplicate to FOA plates, with one plate incubated at 23 °C and the other at 37 °C for 5 days. We isolated 145 temperature-sensitive clones, and we picked 14 for further analysis. We used the *ulp1-333* allele.

Protein purification and enzyme assays. Recombinant proteins were expressed in *E. coli* and purified. GST-Ulp1 was purified on glutathione-agarose, using elution with glutathione. His₆-tagged substrates were purified on a Talon affinity matrix, using elution with 100 mM imidazole. Smt3-Met- β -galactosidase was purified on an aminophenyl-thiopyranogalactoside column. Assays of cleavage of Smt3-fusion proteins were done at 30 °C with different concentrations of substrate and GST-Ulp1 in a reaction mixture containing 150 mM NaCl, 1 mM DTT, and 10 mM Tris-HCl, pH 8.0. For analysis of

SUMO1-ranGAP1 cleavage, transcription and translation of complementary DNA clones were done according to the manufacturer's instructions (Novagen). After 30 min at 30 °C, the reactions were stopped by addition of 10 μ l RNase A (10 mg ml⁻¹). NEM was added to 1 mM, and following a 15-min incubation at 23 °C, β -mercaptoethanol and L-cysteine were added (each to 2 mM) to inactivate any remaining NEM; 5 μ l of this mixture were used in each cleavage reaction.

Cell-cycle analysis. For FACS analysis, log-phase cultures of cells were diluted to 5 $\times$ 10⁷ cells per ml and incubated with 5 μ g ml⁻¹ α -factor in YPD medium for 3 h at 23 °C. Cells were washed twice with YPD, resuspended in prewarmed YPD, and incubated at 37 °C. Cells were fixed in 70% ethanol, digested with RNase A, and stained with propidium iodide. For tubulin immunostaining, log-phase cells were shifted to 37 °C, collected by filtration, and fixed in 50 mM potassium phosphate, pH 6.5, 1 mM MgCl₂ and 4% formaldehyde for 2 h. Fixed cells were washed once, treated with lyticase (ICN), and spotted on polylysine-coated cover slips. Anti- β -tubulin (Boehringer) was the primary antibody, and a rhodamine-conjugated rabbit anti-mouse IgG was the secondary antibody. Following both antibody incubations, the samples were washed eight times with 10 μ l PBS containing 0.15% BSA, 0.1% octyl glucoside and 1% non-fat dry milk. Following air drying, samples were mounted in PBS, 90% glycerol, 0.1% phenylene diamine and 25 ng ml⁻¹ DAPI.

Received 16 December 1998; accepted 14 February 1999.

- Pickart, C. M. Targeting of substrates to the 26S proteasome. *FASEB J.* **11**, 1055-1066 (1997).
- Varshavsky, A. The ubiquitin system. *Trends Biochem. Sci.* **22**, 383-387 (1997).
- Johnson, P. R. & Hochstrasser, M. SUMO-1: ubiquitin gains weight. *Trends Cell Biol.* **7**, 408-413 (1997).
- Saitoh, H., Pu, R. T. & Dasso, M. SUMO-1: wrestling with a new ubiquitin-related modifier. *Trends Biochem. Sci.* **22**, 374-376 (1997).
- Seufert, W., Fitcher, B. & Jentsch, S. A ubiquitin-conjugating enzyme involved in the degradation of both S- and M-phase cyclins. *Nature* **373**, 78-81 (1995).
- Johnson, E. S., Schwenhorst, I., Dohmen, R. J. & Blobel, G. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aop1/Uba2p heterodimer. *EMBO J.* **16**, 5509-5519 (1997).
- Saitoh, H. et al. Ubc9p and the conjugation of SUMO-1 to RanGAP1 and RanBP2. *Curr. Biol.* **8**, 121-124 (1998).
- Müller, S., Matunis, M. J. & Dejean, A. Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. *EMBO J.* **17**, 61-70 (1998).
- Desterro, J. M. P., Rodriguez, M. S. & Hay, R. T. SUMO-1 modification of I κ B α inhibits NF- κ B activation. *Mol. Cell* **2**, 233-239 (1998).
- Wilkinson, K. D. & Hochstrasser, M. in *Ubiquitin and the Biology of the Cell* (eds Peters, J. M., Harris, J. R. & Finley, D.) 99-125 (Plenum, New York, 1998).
- Rose, M. D., Novick, P., Thomas, J. H., Botstein, D. & Fink, G. R. A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. *Gene* **60**, 237-243 (1987).
- Tobias, J. W. & Varshavsky, A. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **266**, 12021-12028 (1991).
- Matunis, M. J., Wu, J. & Blobel, G. SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex. *J. Cell Biol.* **140**, 499-509 (1998).
- Meluh, P. B. & Koshland, D. Evidence that the *MIF2* gene of *Saccharomyces cerevisiae* encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Mol. Biol. Cell* **6**, 793-807 (1995).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).
- Ding, J., McGrath, W. J., Sweet, R. M. & Mangel, W. F. Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO J.* **15**, 1778-1783 (1996).
- López-Otin, C., Simón-Mateo, C., Martínez, L. & Vinuela, E. Gly-Gly-X, a novel consensus sequence for the proteolytic processing of viral and cellular proteins. *J. Biol. Chem.* **264**, 9107-9110 (1989).
- Lee, P. & Hruby, D. E. Proteolytic cleavage of vaccinia virus virion proteins. Mutational analysis of the specificity determinants. *J. Biol. Chem.* **269**, 8616-8622 (1994).
- Schwarz, S. E., Matuschewski, K., Liakopoulos, D., Sheffner, M. & Jentsch, S. The ubiquitin-like proteins SMT3 and SUMO-1 are conjugated by the UBC9 E2 enzyme. *Proc. Natl Acad. Sci. USA* **95**, 560-564 (1998).
- Papa, F. & Hochstrasser, M. The yeast *DOA4* gene encodes a deubiquitinating enzyme related to a product of the human *trc-2* oncogene. *Nature* **366**, 313-319 (1993).
- Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405-439 (1996).
- Deshaies, R. J. Phosphorylation and proteolysis: partners in the regulation of cell division in budding yeast. *Curr. Opin. Genet. Dev.* **7**, 7-16 (1997).
- Kamitani, T. et al. Identification of three major sentrinization sites in PML. *J. Biol. Chem.* **273**, 26675-26682 (1998).
- Müller, S., Miller, W. H. J. & Dejean, A. Trivalent antimonials induce degradation of the PML-RAR oncoprotein and reorganization of the promyelocytic leukemia nuclear bodies in acute promyelocytic leukemia NB4 cells. *Blood* **92**, 4308-4316 (1998).
- Ausubel, F. M. et al. *Current Protocols in Molecular Biology* (Wiley, New York, 1989).
- Johnson, P. R., Swanson, R., Rakhilina, L. & Hochstrasser, M. Degradation signal masking by heterodimerization of MATa2 and MATa1 blocks their mutual destruction by the ubiquitin-proteasome pathway. *Cell* **94**, 217-227 (1998).

Acknowledgements. We thank M. Matunis for the RanGAP1 cDNA expression constructs; R. Swanson for help with immunofluorescence microscopy; R. Cohen for ubiquitin aldehyde; J. Choy for help with FACS analysis; B. Lieberman for plasmids; the University of Chicago Oligopeptide Synthesis facility for N-terminal protein sequencing; and A. Amerik, J. Laney, S. Swaminathan, R. Swanson and A. Turkewitz for critical reading of the manuscript. This work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to M.H. (e-mail: hoc1@midway.uchicago.edu).

The kinase TAK1 can activate the NIK-I κ B as well as the MAP kinase cascade in the IL-1 signalling pathway

Jun Ninomiya-Tsuji*, Kazuya Kishimoto*, Atsushi Hiyama*, Jun-ichiro Inoue†, Zhaodan Cao‡ & Kunihiro Matsumoto*

* Department of Molecular Biology, Graduate School of Science, Nagoya University, and CREST, Japan Science and Technology Corporation, Chikusa-ku, Nagoya 464-01, Japan

† Department of Oncology, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan

‡ Tularik Inc. Two Corporate Drive, South San Francisco, California 94080, USA

Interleukin-1 (IL-1) is a proinflammatory cytokine that has several effects in the inflammation process. When it binds to its cell-surface receptor, IL-1 initiates a signalling cascade that leads to activation of the transcription factor NF- κ B and is relayed through the protein TRAF6 and a succession of kinase enzymes, including NF- κ B-inducing kinase (NIK) and I κ B kinases (IKKs)^{1–7}. However, the molecular mechanism by which NIK is activated is not understood. Here we show that the MAPKK kinase TAK1 (ref. 8) acts upstream of NIK in the IL-1-activated signalling pathway and that TAK1 associates with TRAF6 during IL-1 signalling. Stimulation of TAK1 causes activation of NF- κ B, which is blocked by dominant-negative mutants of NIK, and an inactive TAK1 mutant prevents activation of NF- κ B that is mediated by IL-1 but not by NIK. Activated TAK1 phosphorylates NIK, which stimulates IKK- α activity. Our results indicate that TAK1 links TRAF6 to the NIK-IKK cascade in the IL-1 signalling pathway.

After binding to the cell-surface type-I IL-1 receptor (IL-1RI), IL-1 triggers a cascade of signalling events, including activation of c-

Jun N-terminal kinase (JNK) and of NF- κ B, which upregulates the expression of many proinflammatory genes in the nucleus⁹. NF- κ B is rendered inactive in the cytoplasm by inhibitory proteins called I κ B in unstimulated cells. In response to extracellular stimuli, the I κ B proteins are phosphorylated on specific serine residues and rapidly degraded, leading to the nuclear localization and activation of NF- κ B^{10–12}. A kinase complex consisting of NIK and IKK- α / β is involved in signal-induced phosphorylation of I κ B proteins^{2–7}. When IL-1 binds to IL-1RI, IL-1RI forms a complex with an IL-1-receptor accessory protein, resulting in the recruitment of MyD88 and of the Ser/Thr kinase IRAK to the receptor. IRAK then dissociates from the receptor complex and interacts with TRAF6, which transduces the IL-1 signal to the NIK-IKK- α / β kinase pathway^{13,14}. TRAF6 has been implicated in the activation of both JNK and NF- κ B¹⁵. In contrast, NIK is essential for IL-1-mediated NF- κ B activation but has no effect on the activation of JNK^{15,16}. Thus, the molecular mechanism linking TRAF6 to the activation of NIK is likely to involve additional signalling molecules. TAK1 is a recently identified MAP kinase kinase kinase (MAP3K) that can activate JNK^{8,17,18}; another protein, TAB1, functions as an activator of TAK1 (ref. 19). We now examine the role of TAK1 in the NIK-IKK cascade that operates in the IL-1 signalling pathway.

Treatment of 293IL-1RI cells with IL-1 resulted in activation of endogenous TAK1 in a time-dependent manner (Fig. 1a). To determine whether TAK1 participates in IL-1 signalling, we investigated the interaction of TAK1 and TAB1 with TRAF6. 293IL-1RI cells were either treated with IL-1 or left untreated; lysates were immunoprecipitated with antibody against TRAF6 and the co-precipitated TAK1 and TAB1 were detected by immunoblot analysis (Fig. 1b). Endogenous TAK1 and TAB1 were found to associate with TRAF6 in IL-1-treated cells but not in untreated cells. Thus, the association of endogenous TRAF6 with TAK1 and TAB1 is ligand-dependent. Next, 293 cells were transiently transfected with an expression vector encoding a Flag-epitope-tagged TRAF6 (Flag-TRAF6) protein. Immunoprecipitation of Flag-TRAF6 caused co-precipitation with TAK1 and TAB1 even from lysates of cells not treated with IL-1 (Fig. 1c). This indicates that TRAF6 constitutively associates with TAB1 and TAK1 when TRAF6 proteins are over-

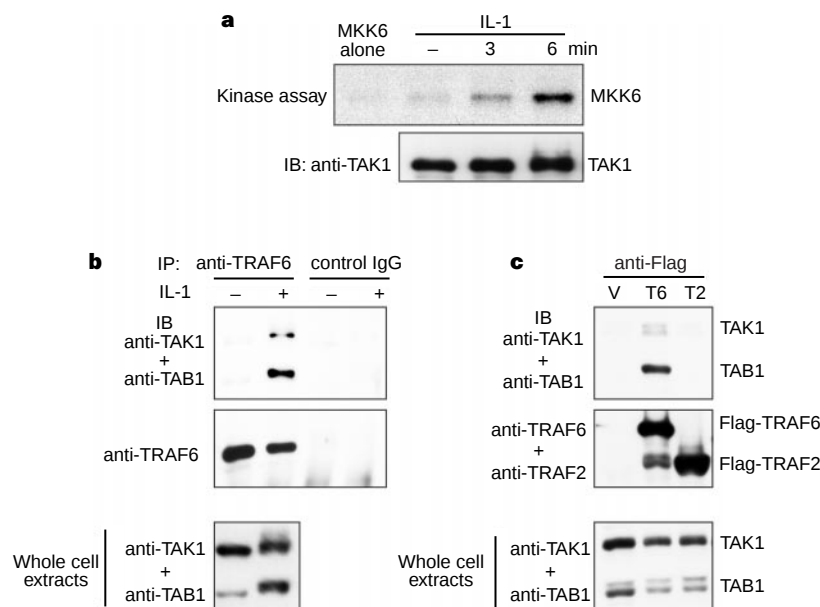


Figure 1 Effect of IL-1 on TAK1. **a**, IL-1-induced activation of TAK1 activity. 293IL-1RI cells were treated with IL-1 (10 ng ml^{–1}). Immunoprecipitated complexes with anti-TAK1 were subjected to kinase reactions with MKK6 and immunoblotted (IB) with anti-TAK1. **b**, **c**, Association of TRAF6 with TAK1. 293IL-1RI cells were stimulated for 10 min with IL-1 or left untreated. Cell lysates were immunopreci-

pitated (IP) with anti-TRAF6 or IgG (**b**). 293 cells were transfected with empty vector (V), Flag-TRAF6 (T6), or Flag-TRAF2 (T2). Cell lysates were immunoprecipitated with anti-Flag (**c**). The immunoprecipitates were immunoblotted with each antibody. Expression of TAK1 and TAB1 was monitored.

expressed and is consistent with the observation that TRAF6 overexpression can activate NF- κ B in the absence of ligand¹. We also investigated the interaction of TAK1 and TAB1 with another member of the TRAF family, TRAF2, which has been implicated in the TNF-mediated signalling pathway¹⁵. In contrast to TRAF6, TAB1-TAK1 did not co-precipitate with TRAF2 complex (Fig. 1c).

TRAF6 has been implicated in IL-1-induced NF- κ B activation¹. To determine whether TAK1 also functions in the activation of NF- κ B, we tested the effects of activation of TAK1 on the expression of an NF- κ B-dependent reporter gene in transiently transfected 293 cells. Co-expression of TAB1 and TAK1 activated the kinase activity of TAK1 and increased transcription of the reporter gene (Fig. 2a). Overexpression of another member of the MAP3K family, MTK1 (ref. 20), failed to induce NF- κ B activation (data not shown). NF- κ B was not activated when TAB1 and a kinase-inactive mutant of TAK1, TAK1(K63W), were co-expressed, indicating that TAK1-mediated NF- κ B activation is dependent on its kinase activity. Activation of NF- κ B by IL-1 requires the successive action of NIK and IKK- α/β ²⁻⁷. To determine the signalling pathways that couple TAK1 to NF- κ B activation, we tested whether dominant-negative forms of NIK could block signalling from TAB1-TAK1. The inactive NIK mutant NIK(KK429-430AA) behaves as a dominant-negative inhibitor of IL-1-induced NF- κ B activation². Expression of NIK(KK429-430AA) blocked the activation of NF- κ B-dependent gene expression by TAB1-TAK1 (Fig. 2b). Furthermore, in contrast to a recent observation²¹, another dominant-negative form, NIK(624-947), also inhibited TAB1-TAK1-induced activation of NF- κ B (Fig. 2b). Overexpression of the inactive TAK1(K63W) mutant inhibited IL-1-induced activation of NF- κ B in a dose-dependent manner, but had little effect on NIK-induced activation of NF- κ B (Fig. 2b). Taken together, these results indicate that

TAB1-TAK1 may function upstream of the NIK-IKK cascade to mediate NF- κ B activation by IL-1.

In addition to NF- κ B activation, IL-1 induces activation of JNK⁹. To test whether TAK1 is involved in IL-1-induced JNK activation, we did JNK assays by using a glutathione S-transferase fusion protein (GST-c-Jun) as substrate. Overexpression of TAK1(K63W) blocked IL-1-induced JNK activation (Fig. 2c), suggesting that TAK1 participates in both NF- κ B and JNK activation by IL-1. In contrast, the expression of NIK(KK429-430AA) at levels that gave maximal inhibition of NF- κ B activation did not impair the ability of either IL-1 or TAB1-TAK1 to activate JNK (Fig. 2c). Thus, bifurcation of the IL-1-induced JNK and NF- κ B activation pathways occurs at TAK1.

Our results raised the possibility that TAK1 could function as the kinase that activates NIK in the IL-1 signalling pathway. We therefore investigated whether activation of TAK1 induces phosphorylation of NIK. Immunoblot analysis of NIK from cells overexpressing wild-type NIK revealed that NIK migrated as a doublet (Fig. 4b, lane 6), possibly as a result of autophosphorylation and/or phosphorylation by activated IKK- α (ref. 3). When expressed alone, NIK(KK429-430AA) also appeared as a doublet, but with a minor form that migrated more slowly than the major form (Fig. 3a, lane 1). IL-1 treatment caused an accumulation of the more slowly migrating form of NIK (Fig. 3a, lane 2). This slower band was eliminated by phosphatase treatment (Fig. 3a, lane 7), suggesting that it represents phosphorylated NIK. When TAK1 was activated by co-expression with TAB1, the amount of slower NIK(KK429-430AA) increased (Fig. 3a, lane 5). This modification was not seen in 293 cells transfected with MTK1 MAP3K (data not shown), indicating that NIK modification is specific for TAK1. Furthermore, expression of the dominant-negative mutant TAK1(K63W) inhib-

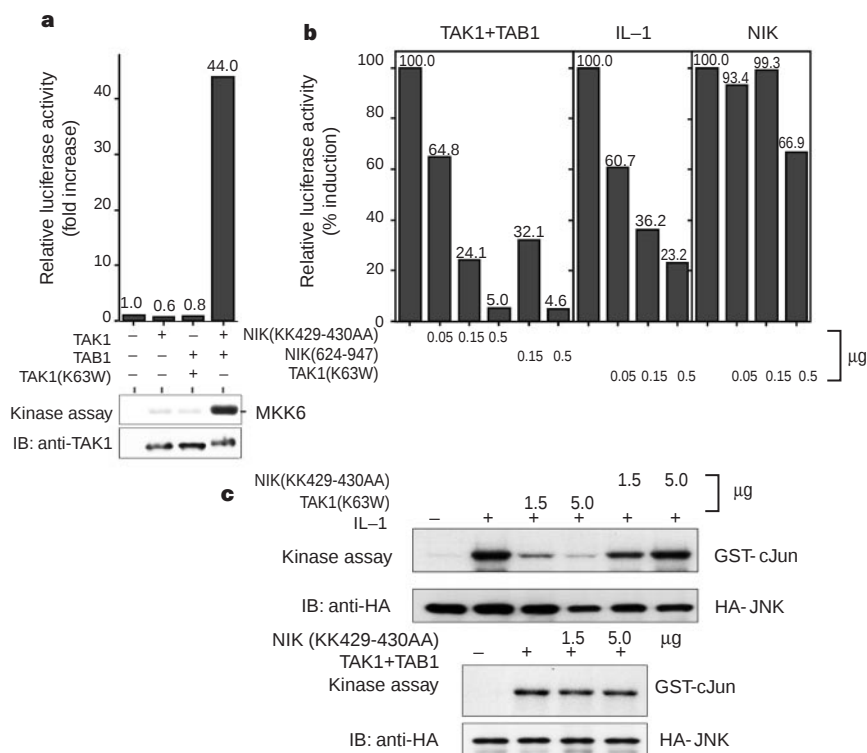


Figure 2 Effect of TAK1 on NF- κ B and JNK activation. **a, b**, Effect of TAK1 on NF- κ B activation. 293 cells were transfected with HA-TAK1, HA-TAK1(K63W) and TAB1, as indicated. Immunoprecipitated complexes with anti-HA were subjected to kinase reactions and immunoblotted with anti-TAK1 (**a**, lower panel). Ig- κ -luciferase plasmid was transfected into 293 cells together with expression vectors, as indicated. Cells were either treated with IL-1 (10 ng ml⁻¹) or left

untreated. Relative luciferase activity was measured. Data are expressed as the fold increase (**a**) or the percentage induction (**b**). **c**, Effect of TAK1 on JNK activation. HA-JNK1 was transfected into 293IL-1R1 cells together with expression vectors, as indicated. Cells were either treated with IL-1 for 10 min or left untreated. Immunoprecipitated complexes with anti-HA were subjected to kinase reactions with GST-cJun and immunoblotted with anti-HA.

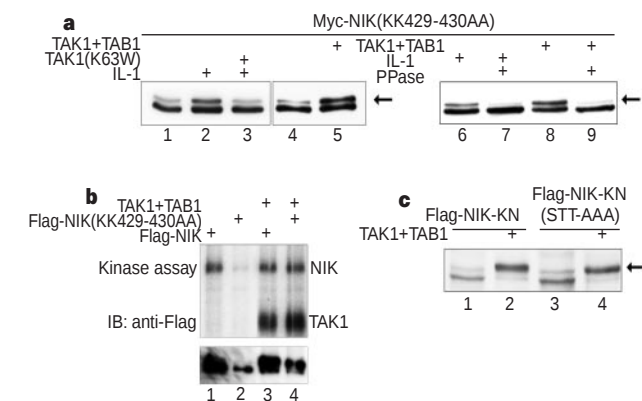


Figure 3 Phosphorylation of NIK. **a, c**, Effect of TAK1 on phosphorylation of NIK. 293 cells were transfected with Myc-NIK(KK429-430AA) (**a**), and Flag-NIK(KK429-430AA) (Flag-NIK-KN) or Flag-NIK(KK429-430AA;STT549,552,559AAA) [Flag-NIK-KN(STT-AAA)] (**c**). Cells were also transfected with TAK1 and TAB1 or TAK1(K63W), as indicated. Cells were treated for 30 min with IL-1 (10 ng ml⁻¹) or left untreated. Cell lysates were incubated with protein phosphatase (PPase). Anti-Myc (**a**) or anti-Flag (**c**) was used to detect NIK in total lysates. Arrows indicate the slowly migrating form of NIK. **b**, *In vitro* phosphorylation of NIK by TAK1. 293 cells were transfected with expression vectors as indicated. Complexes immunoprecipitated with anti-Flag were incubated with [γ -³²P]ATP and analysed by autoradiography. The migration positions of NIK and TAK1 are indicated.

ited IL-1-induced modification of NIK (Fig. 3a, lane 3). These findings indicate that IL-1 signalling activates TAK1, causing phosphorylation of NIK.

As activation of TAK1 resulted in NIK phosphorylation, we investigated whether TAK1 can phosphorylate NIK *in vitro* (Fig. 3b). We transiently expressed Flag-epitope-tagged wild-type NIK (Flag-NIK), or the inactive mutant Flag-NIK(KK429-430AA) in 293 cells. Epitope-tagged proteins were immunoprecipitated with anti-Flag antibody, and incubated with [γ -³²P]ATP. In these assays, wild-type NIK was autophosphorylated (Fig. 3b, lane 1), whereas NIK(KK429-430AA) was not (Fig. 3b, lane 2). When TAK1 and TAB1 were co-expressed with Flag-NIK(KK429-430AA), activated TAK1 phosphorylated NIK(KK429-430AA) and it became auto-phosphorylated *in vitro* (Fig. 3b, lane 4).

Phosphorylation of Ser and/or Thr residues in the kinase-activation loop (between subdomains VII and VIII) is essential for activation of many protein kinases²². There is a serine (residue 549) and two threonines (residues 552 and 559) in the activation loop of NIK (ref. 2). We generated a mutant NIK, NIK(STT549,552,559AAA), in which Ser 549, Thr 552 and Thr 559 were replaced with alanine residues. In agreement with recent results²³, NIK(STT549,552,559AAA) prevented NIK from activating NF- κ B and functioned as a dominant-negative inhibitor of IL-1-mediated activation of NF- κ B, like the NIK(KK429-430AA) mutant (data not shown). Thus, residues in the activation loop of NIK are critical for the regulation of NIK. To test whether TAK1 participates in the phosphorylation of residues within the activation loop of NIK, we generated a Flag-epitope-tagged NIK mutant, Flag-NIK(KK429-430AA;STT549,552,559AAA), in the background of the inactive NIK(KK429-430AA) mutation in order to eliminate NIK auto-phosphorylation activity (Fig. 3c). When TAK1 was activated by co-expression with TAB1, the NIK(KK429-430AA;STT549,552,559AAA) mutant form was still modified (Fig. 3c, lane 4), indicating that TAK1 can induce phosphorylation of NIK residues apart from the Ser 549, Thr 552 and Thr 559 residues in the activation loop.

To determine whether TAK1 associates with NIK, we co-transfected TAK1 and Flag-NIK(KK429-430AA) into 293 cells and then immunoprecipitated the lysates (Fig. 4a). TAK1 was detected in NIK(KK429-430AA) immunoprecipitates (Fig. 4a, lane 2). We also

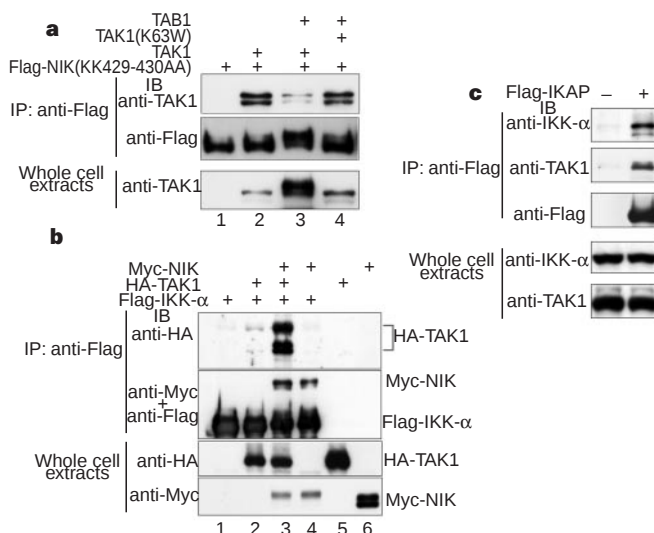
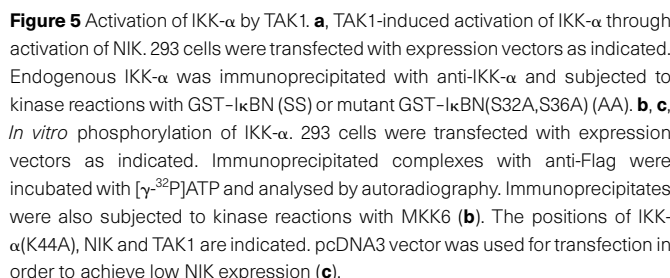


Figure 4 Interaction of TAK1 with components of the IKK complex. 293 cells were transfected with expression vectors as indicated. Cell lysates were immunoprecipitated with anti-Flag. **a–c**, Immunoprecipitates were immunoblotted with anti-TAK1 and anti-Flag (**a**), anti-HA, anti-Myc and anti-Flag (**b**), and anti-IKK- α , anti-TAK1 and anti-Flag (**c**). Immunoprecipitated HA-TAK1 presented as multiple bands, probably generated by phosphorylation by NIK and/or IKK- α (**b**). The expression of TAK1 (**a**), HA-TAK1 and Myc-NIK (**b**), and of IKK- α and TAK1 (**c**) was monitored.

did the same experiment using cells in which TAB1 was co-expressed as well in order to activate TAK1. Although co-expression of NIK(KK429-430AA), TAB1 and TAK1 caused phosphorylation of NIK(KK429-430AA), the association of NIK(KK429-430AA) and TAK1 decreased compared with cells containing non-activated TAK1 (Fig. 4a, lane 3). Thus, the inactive form of TAK1 preferentially associates with NIK(KK429-430AA). Alternatively, TAB1 competes with NIK for binding to TAK1. To exclude this possibility, we examined the interaction between NIK and inactive TAK1(K63W) in the presence of TAB1. TAK1(K63W) associated with NIK(KK429-430AA) more efficiently than did wild-type TAK1 (Fig. 4a, lane 4). Wild-type and inactive TAK1 associated with TAB1 to the same extent (data not shown). These results indicate that, once activated, TAK1 phosphorylates NIK and it may then be released from the NIK–IKK complex.

We next investigated whether TAK1, NIK and IKK- α are present in the same complex by using co-transfection and co-immunoprecipitation methods. Flag-epitope-tagged IKK- α (Flag-IKK- α) was transiently co-expressed in 293 cells together with either Myc-epitope-tagged NIK (Myc-NIK) or haemagglutinin-epitope-tagged TAK1 (HA-TAK1). Cell lysates were immunoprecipitated with anti-Flag antibody and analysed by immunoblotting using anti-Myc or anti-HA antibodies to detect NIK or TAK1, respectively (Fig. 4b). In this assay, NIK co-precipitated with IKK- α (Fig. 4b, lane 4), whereas TAK1 interacted only weakly with IKK- α (Fig. 4b, lane 2). To determine whether NIK can mediate interaction between TAK1 and IKK- α , we co-expressed the three proteins in 293 cells. There was a strong interaction between TAK1 and IKK- α (Fig. 4b, lane 3), indicating that TAK1 interacts with a complex consisting of NIK and IKK. Consistent with this, we found that TAK1 also associated with IKAP (IKK-complex-associated protein) (Fig. 4c). IKAP acts as a scaffold that assembles IKKs with NIK²⁴. Flag-epitope-tagged IKAP (Flag-IKAP) was transfected into 293 cells and the lysates were then immunoprecipitated. Endogenous TAK1 and IKK- α were detected in IKAP immunoprecipitates.

These results indicate that activated TAK1 stimulates the kinase activity of NIK by phosphorylation. To test whether TAK1 stimu-



NIK is a MAP3K-related kinase that phosphorylates Ser 176 during the activation of IKK- α (ref. 25). MAP3K enzymes activate MAP2K enzymes by phosphorylating serine and threonine residues in the activation loop. TAK might therefore phosphorylate IKK, so we co-expressed the tagged inactive IKK- α mutant Flag-IKK- α (K44A) with TAK1 and TAB1 or NIK (Fig. 5b), immunoprecipitated the epitope-tagged proteins and incubated them with [γ - 32 P]ATP. NIK phosphorylated IKK- α (Fig. 5b, lane 2) as expected²⁵, whereas TAK1 barely phosphorylated IKK- α (Fig. 5b,

Figure 6 Model for the role of TAK1 in the IL-1 signalling pathway. See text for details

lane 4). When incubated with bacterially expressed MKK6, TAK1 but not NIK phosphorylated MKK6: thus, TAK1 and NIK have different substrate specificities. When we co-transfected Flag-IKK- α (K44A), TAK1 and TAB1 with small amounts of NIK, and measured the phosphorylation of IKK- α (K44A), we found that TAK1 stimulated the phosphorylation of IKK- α (K44A) by NIK (Fig. 5c). Taken together, these results indicate that activated TAK1 promotes the phosphorylation of NIK and hence activation of IKK by NIK.

Our results identify TAK1 as a kinase that links TRAF6 to the NIK–IKK cascade in the IL-1 signalling pathway (Fig. 6). TRAF6 also interacts with IRAK and NIK (refs 1, 15), so it may serve as a docking site for these signalling molecules in the pathway. TAK1 is recruited to TRAF6 in response to IL-1 stimulation, which may activate TAK1 and thus NIK. It is still unclear how TAK1 is activated: additional components may be needed. NIK shares homology with MAP3K (ref. 2). It would be rather unusual for TAK1 MAP3K to phosphorylate another MAP3K-like kinase like NIK, although there is no direct evidence yet that NIK functions as a MAP3K. Although NIK phosphorylates IKK- α *in vitro*²⁵, IKK- α is not homologous to MAP2K (refs 3–5). Furthermore, NIK is not involved in activation of JNK or p38 MAP kinases^{15,16}. Therefore, TAK1 MAP3K may phosphorylate both MAP2K and the MAP3K-like kinase NIK (Fig. 6). A conserved threonine residue, Thr 559, in NIK that is important for its catalytic activity²³ is not phosphorylated by TAK1 but is autophosphorylated²³. Our preliminary results indicate that TAK1 may phosphorylate the N-terminal region of NIK (data not shown), its non-catalytic domain, which might partially activate NIK so that it can autophosphorylate in the catalytic domain, possibly at Thr 559, and stimulate catalytic activity. Our results should enhance understanding of the IL-1 signalling pathway.

Expression vectors and antibodies. To overexpress TAK1(K63W), we constructed the expression vector pCMV-TAK1(K63W), encoding TAK1(K63W) under the control of the cytomegalovirus (CMV) promoter. Mammalian expression vectors encoding Myc-NIK, Myc-NIK(KK429-4309AA), Flag-IKK- α , Flag-IKK- α (K44A), HA-JNK, TAK1, TAB1 and HA-TAK1 have been described^{3,8,17}. The NIK mutant, NIK(STT549,552, 559AAA), was generated by overlapping PCR. Polyclonal rabbit antibodies against TAK1 and TAB1 were produced against peptides corresponding to amino acids 554–579 of TAK1 and amino acids 480–500 of TAB1, respectively.

NF- κ B-dependent reporter assays. 293 cells (1.6×10^5 cells per well) were

seeded into 6-well (35-mm) plates. Cells were transfected at 2 days after seeding by the calcium phosphate precipitation method with an Ig- κ -luciferase reporter gene plasmid and each expression vector. The total DNA concentration (1.7 μ g) was kept constant by supplementing with empty vector. Luciferase activity was determined using the Luciferase Assay System (Promega). A β -galactosidase vector (0.1 μ g) under the control of the β -actin promoter was used for normalizing transfection efficiencies.

Protein phosphatase treatment. Cells were lysed in 1% NP-40 buffer containing 10 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA and 1 mM PMSF without phosphatase inhibitors. Aliquots of lysates were incubated with or without Lambda Protein Phosphatase (New England Biolabs) in phosphatase buffer containing 50 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 5 mM DTT, 0.01% Brij-35 and 2 mM MnCl₂ at 30 °C for 30 min.

In vitro phosphorylation assays. To construct GST-I κ BN, a cDNA encoding the first 72 amino acids of human I κ B- α was subcloned into pGEX2T (Pharmacia). To construct GST-I κ BN(S32A,S36A), Ser32 and Ser36 were changed to alanine by site-directed mutagenesis. Endogenous IKK- α was immunoprecipitated from 293 cells. Aliquots of immunoprecipitates were incubated with 5 μ g bacterially expressed GST-I κ BN or GST-I κ BN(S32A,S36A) proteins in 15 μ l kinase buffer containing 20 mM Tris-HCl (pH 7.6), 20 mM MgCl₂, 20 mM β -glycerophosphate, 1 mM EDTA, 1 mM sodium orthovanadate, 0.4 mM PMSF, 1 mM ATP, 20 mM creatine phosphate and 5 μ Ci [γ -³²P]ATP at 37 °C for 30 min. Samples were resolved by SDS-PAGE, and phosphorylated GST-I κ BN was visualized by autoradiography.

Received 24 November 1998; accepted 25 January 1999.

1. Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. & Goeddel, D. V. TRAF6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
2. Malinin, N. L., Boldin, M. P., Kovalenko, A. V. & Wallach, D. MAP3K-related kinase involved in NF- κ B induction by TNF, CD95 and IL-1. *Nature* **385**, 540–544 (1997).
3. Regnier, C. H. et al. Identification and characterization of an I κ B kinase. *Cell* **90**, 373–383 (1997).
4. DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive I κ B kinase that activates the transcription factor NF- κ B. *Nature* **388**, 548–554 (1997).
5. Mercurio, F. et al. IKK-1 and IKK-2: cytokine-activated I κ B kinases essential for NF- κ B activation. *Science* **278**, 860–866 (1997).
6. Woronicz, J. D., Gao, X., Cao, Z., Rothe, M. & Goeddel, D. V. I κ B kinase- β : NF- κ B activation and complex formation with I κ B kinase- α and NIK. *Science* **278**, 866–869 (1997).
7. Zandi, E., Rothwarf, D. M., Delhase, M., Hayakawa, M. & Karin, M. The I κ B kinase complex (IKK) contains two kinase subunits, IKK α and IKK β , necessary for I κ B phosphorylation and NF- κ B activation. *Cell* **91**, 243–252 (1997).
8. Yamaguchi, K. et al. Identification of a member of the MAPKKK family as a potential mediator of TGF- β signal transduction. *Science* **270**, 2008–2011 (1995).
9. Dinarello, C. A. Biologic basis for interleukin-1 in disease. *Blood* **87**, 2095–2147 (1996).
10. Thanos, D. & Maniatis, T. NF- κ B: a lesson in family values. *Cell* **80**, 529–532 (1995).
11. Verma, I. M., Stevenson, J. K., Schwarz, E. M., van Antwerp D. & Miyamoto, S. Rel/NF- κ B/I κ B family: intimate tales of association and dissociation. *Genes Dev.* **9**, 2723–2735 (1995).
12. Baeuerle, P. A. & Baltimore, D. NF- κ B: ten years after. *Cell* **87**, 13–20 (1996).
13. Cao, Z., Henzel, W. J. & Gao, X. IRAK: a kinase associated with the interleukin-1 receptor. *Science* **271**, 1128–1131 (1996).
14. Wesche, H., Henzel, W. J., Shillinglaw, W., Li, S. & Cao, Z. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity* **7**, 837–847 (1997).
15. Song, H. Y., Regnier, C. H., Kirschning, C. J., Goeddel, D. V. & Rothe, M. Tumor necrosis factor (TNF)-mediated kinase cascades: bifurcation of nuclear factor- κ B and c-Jun N-terminal kinase (JNK/SAPK) pathways at TNF receptor-associated factor 2. *Proc. Natl Acad. Sci. USA* **94**, 9792–9796 (1997).
16. Karin, M. & Delhase, M. JNK or IKK, AP-1 or NF- κ B: which are the targets for MEK kinase 1 action? *Proc. Natl Acad. Sci. USA* **95**, 9067–9069 (1998).
17. Shirakabe, K. et al. TAK1 mediates the ceramide signaling to stress-activated protein kinase/c-Jun N-terminal kinase. *J. Biol. Chem.* **272**, 8141–8144 (1997).
18. Moriguchi, T. et al. A novel kinase cascade mediated by mitogen-activated protein kinase kinase 6 and MKK3. *J. Biol. Chem.* **271**, 13675–13679 (1996).
19. Shibuya, H. et al. TAB1: an activator of the TAK1 MAPKKK in TGF- β signal transduction. *Science* **272**, 1179–1182 (1996).
20. Takekawa, M., Posas, F. & Saito, H. A human homolog of the yeast Ssk2/Ssk22 MAP kinase kinase kinases, MTK1, mediates stress-induced activation of the p38 and JNK pathways. *EMBO J.* **16**, 4973–4982 (1997).
21. Sakurai, H., Shigemori, N., Hasegawa, K. & Sugita, T. TGF- β -activating kinase 1 stimulates NF- κ B activation by an NF- κ B-inducing kinase-independent mechanism. *Biochem. Biophys. Res. Commun.* **243**, 545–549 (1998).
22. Johnson, L. N., Noble, M. E. M. & Owen, D. J. Active and inactive protein kinases: structural basis for regulation. *Cell* **85**, 149–158 (1996).
23. Lin, X. et al. Molecular determinants of NF- κ B-inducing kinase action. *Mol. Cell. Biol.* **18**, 5899–5907 (1998).
24. Cohen, L., Henzel, W. J. & Baeuerle, P. A. IKAP is a scaffold protein of the I κ B kinase complex. *Nature* **395**, 292–296 (1998).
25. Ling, L., Cao, Z. & Goeddel, D. V. NF- κ B-inducing kinase activates IKK- α by phosphorylation of Ser-176. *Proc. Natl Acad. Sci. USA* **95**, 3792–3797 (1998).

Acknowledgements. We thank D. Goeddel for advice and for discussion; P. Baeuerle, T. Fujita, E. Nishida, H. Saito, H. Sakurai, H. Shibuya and D. Wallach for materials; and M. Lamphier and R. Ruggieri for critically reading the manuscript. Supported by special grants for CREST, Advanced Research on Cancer, from the Ministry of Education, Culture and Science of Japan, and by HFSP (K.M.).

Correspondence and requests for materials should be addressed to K.M. (e-mail: g44177a@nucc.cc.nagoya-u.ac.jp).

NF-AT activation requires suppression of Crm1-dependent export by calcineurin

Jiangyu Zhu & Frank McKeon

Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Nuclear import of the NF-AT transcription factors during T-cell activation requires the calcium-activated phosphatase calcineurin, which unmasks nuclear-location signals on NF-AT (refs 1–5). We show here that the nuclear import of NF-ATs is not sufficient to activate NF-AT target genes, as NF-ATs are subject to a futile cycling across the nuclear envelope owing to engagement with the exportin protein Crm1 (refs 6–8). Calcineurin suppresses this futile cycling by a non-catalytic mechanism involving the masking of nuclear export signals on NF-AT targeted by Crm1. This clustering of binding sites for calcineurin and Crm1 on NF-AT establishes an inherent competition between these molecules that imparts exquisite calcium sensitivity to the shuttling dynamics of the NF-AT transcription factors. Such a balance between nuclear import and export may regulate the action of other transcription factors.

The fungal metabolite leptomycin B is a specific inhibitor of the nuclear export signal receptor Crm1 (refs 6–12). Leptomycin B also inhibited NF-AT nuclear export following calcium ionophore-induced nuclear import (see Figs A–C in Supplementary Information). To assess the role of Crm1 in NF-AT4 shuttling dynamics, we investigated whether Crm1 overexpression would prevent NF-AT from moving to the nucleus. Overexpressed Crm1 had no effect on the cytoplasmic localization of NF-AT4 in resting cells, but completely blocked the nuclear localization of NF-AT4 in response to calcium ionophore (see Figs E, F in Supplementary Information). Similar results were obtained with NF-AT4(N), which lacks the Rel homology domain (see Figs D, F in Supplementary Information). We have shown previously that the nuclear-localization signal (NLS) on NF-AT is masked by the NLS-masking, or Z, domain², and that the NF-AT4(N) Δ Z mutant is constitutively nuclear (Fig. 1a). Significantly, overexpressed Crm1 efficiently exported NF-AT4(N) Δ Z from the nucleus (Fig. 1a), indicating that Crm1 actively exports NF-AT from the nucleus rather than interferes with its intramolecular NLS-masking mechanism.

To define regions of NF-AT4 required for Crm1-mediated export, we screened NF-AT(N) Δ Z mutants for those that could bypass Crm1 and remain in the nucleus. One mutant that lacked the amino terminus of NF-AT(N) Δ Z was exclusively nuclear, despite Crm1 overexpression (Fig. 1b). Two subdomains of this amino-terminal region were found to be sufficient for nuclear export of NF-AT4(N) Δ Z by Crm1 (Fig. 1b). We called these regions NES1 (amino acids 31–96) and NES2 (amino acids 99–154) to reflect the fact that each contained an autonomous NES sequence targeted by Crm1.

Crm1 has been shown to bind the nuclear-export signal (NES) of the Rev protein of human immunodeficiency virus (HIV)⁷. To determine whether the functionally defined NES sequences on NF-AT4 physically bind Crm1, we assayed interactions between the NF-AT mutants and Crm1. Crm1 displayed good affinity for a fusion protein consisting of glutathione S-transferase (GST) and NF-AT4(N), with no significant binding to GST (Fig. 1c). Deletion of the NES1 domain diminished, but did not abolish, Crm1 association. Similarly, a GST–NF-AT4(N) mutant lacking the

NES2 domain also bound Crm1, but less than the wild-type fusion protein (not shown). However, the deletion of both the NES1 and NES2 domains completely abolished the interaction between Crm1 and NF-AT4(N) (Fig. 1c). These data indicate that the NESs identified by functional analyses also mediate interactions between Crm1 and NF-AT4, and that each of these NESs is sufficient for Crm1 binding and Crm1-mediated nuclear export.

To demonstrate further that the NESs identified in NF-AT4 are essential for the nuclear export of NF-AT4, we used a heterokaryon cell fusion assay for NES activity¹³ in which BHK cells expressing NF-AT4(N) Δ Z mutants were fused to HeLa cells expressing an NES-GFP-NLS shuttling protein (where GFP is green fluorescent protein). As a positive control for nuclear export and for identifying heterokaryons, we used the NES-GFP-NLS fusion protein to label both the large, ellipsoidal HeLa nuclei and the smaller, spherical nuclei of BHK cells (Fig. 1d). NF-AT4(N) Δ Z mutants with NES activity should be exported to the common cytoplasm of the heterokaryon, and then be imported to both the HeLa and BHK nuclei. We examined the behaviour of various deletion mutants made in NF-AT4(N) Δ Z and found that NF-AT4(N) Δ Z/ Δ NES1, which retains one potential NES (NES2), was transported from the BHK to the HeLa nucleus (Fig. 1d). Similar results were obtained with NF-AT4(N) Δ Z/ Δ NES2 (not shown). In contrast, the NF-AT4(N) Δ Z/ Δ NES1/ Δ NES2 mutant, lacking both putative NESs, remained exclusively in the BHK nuclei (Fig. 1d). Thus, in heterokaryons, the NF-AT4(N) Δ Z/ Δ NES1/ Δ NES2 mutant could not be exported from the BHK nucleus, indicating that this mutant had lost all functional NES sequences.

Activated calcineurin has been shown to bind directly the regulatory, N-terminal half of NF-AT4 (refs 1, 14, 15). This calcineurin-binding region of NF-AT has been narrowed down to a

27-amino-acid sequence (amino acids 100–126), as deletion of this sequence renders NF-AT4 cytoplasmic despite ionomycin treatment². Further, a peptide corresponding to this sequence interferes with calcineurin–NF-AT interactions¹⁶. Significantly, this calcineurin-binding site lies within the NES2 domain, and the NF-AT4(N) mutant lacking NES2 failed to translocate to the nucleus in response to ionomycin (Fig. 2a). These data indicate that the NES2 sequence functionally overlaps that of a calcineurin-binding site on NF-AT4. However, we noted that the NF-AT4(N) Δ NES2 mutant could be driven to the nucleus by overexpression of the constitutively active calcineurin Δ CnA, indicating that additional, although weaker, calcineurin-binding sites exist on NF-AT4 (Fig. 2a)². We then sought to identify these additional sites by co-expressing Δ CnA and various deletion mutants of NF-AT4(N) in BHK cells. As Fig. 2a shows, deletion of the regions encompassing both NES1 and NES2 in NF-AT4(N) resulted in a total loss of response not only to calcium ionophore, but also to overexpressed Δ CnA, suggesting that NES1 and NES2 together comprise the sites necessary for NF-AT interaction with calcineurin.

We examined the interactions between NF-AT4(N) mutants and Δ CnA further in a co-immunoprecipitation assay. The deletion of either NES1 or NES2 diminished the association of Δ CnA (Fig. 2b) and deletion of both NES1 and NES2 completely abolished the interaction between NF-AT4 and calcineurin (Fig. 2b). Thus, the regions implicated in calcineurin interactions are essentially contained within the NES1 and NES2 domains also required for Crm1-mediated nuclear export. As the NES1 and NES2 domains defined

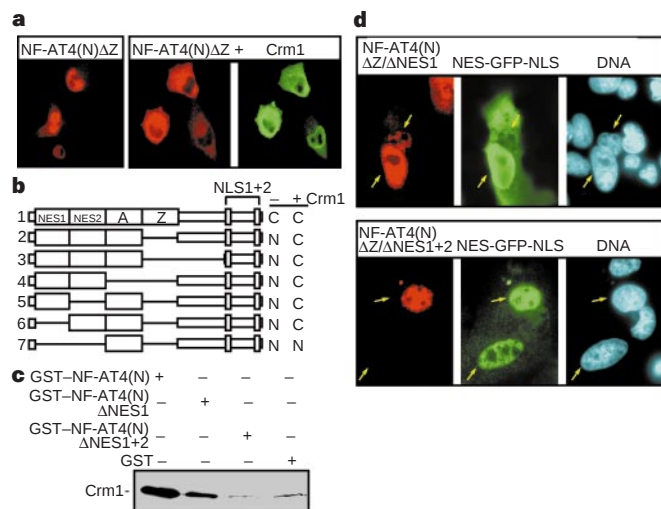


Figure 1 Nuclear export signals and Crm1-binding sites in NF-AT4. **a**, Immunofluorescence of the NF-AT4(N) Δ Z mutant, which lacks the NLS-masking domain, in the absence (left panel) or presence (right panels) of Crm1. **b**, Various deletion mutants of NF-AT4(N) were expressed with or without Crm1 in BHK cells and detected by immunofluorescence. NES1, amino acids 31–96; NES2, amino acids 99–154. C, predominantly cytoplasmic; N, predominantly nuclear. **c**, CRM1 binds the NES1 and NES2 domains of NF-AT4(N). GST-NF-AT4(N) deletion mutant fusion proteins were incubated with lysates of BHK cells expressing HA-tagged Crm1, and bound Crm1 was detected by immunoblotting with the anti-HA antibody. **d**, Heterokaryon assay for nuclear-export-signal activity of NF-AT4 deletion mutants in **b**. HeLa cells expressing NES-GFP-NLS were fused with BHK cells expressing Myc-tagged NF-AT4(N). Fused cells were identified by GFP and anti-Myc immunofluorescence staining. Images for mutant 6 (Δ NES1) and mutant 7 (Δ NES1+2) are shown. The HeLa and BHK nuclei in the fused cells are indicated by arrows.

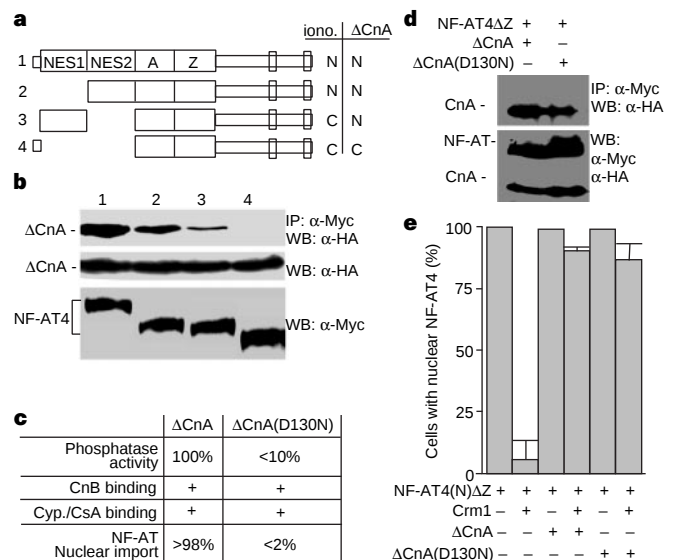


Figure 2 Calcineurin binding abolishes NF-AT4 nuclear export by Crm1. **a**, The calcineurin-binding sites in NF-AT4 overlap nuclear export signals. Subcellular localization of NF-AT4(N) deletion mutants in BHK cells treated with calcium ionophore (iono.). Alternatively, the localization of the deletion mutants was determined on co-expression with constitutively active calcineurin (Δ CnA). C, predominantly cytoplasmic; N, predominantly nuclear. **b**, NF-AT4(N) mutants (as in **a**) were co-expressed with HA-tagged Δ CnA. Mutants were immunoprecipitated and associated Δ CnA was examined by immunoblotting. IP, immunoprecipitation; WB, western blotting. **c**, Functional characterization of phosphatase-impaired calcineurin mutant, Δ CnA(D130N). Phosphatase assays, as well as CnB and cyclosporin A/cyclophilin complex binding, were done as previously described²³. **d**, Co-immunoprecipitation of Δ CnA(D130N) with NF-AT4. NF-AT4 Δ Z was co-expressed with HA- Δ CnA or HA- Δ CnA(D130N), immunoprecipitated and associated Δ CnA detected by immunoblotting with an anti-HA antibody. **e**, Calcineurin binding abolishes NF-AT4(N) Δ Z export by Crm1. GFP-NF-AT4(N) Δ Z was co-transfected with various combinations of Crm1, Δ CnA, and Δ CnA(D130N), and the percentage of cells showing predominantly nuclear GFP was determined.

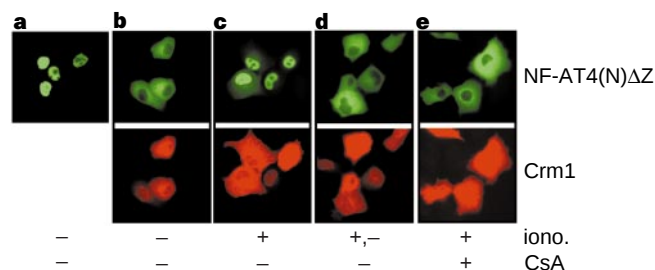


Figure 3 Calcium regulates nuclear export by calcineurin. **a**, GFP-tagged NF-AT4(N)ΔZ mutant localized to nucleus of BHK cells. **b**, GRP-NF-AT4(N)ΔZ (green) is cytoplasmic when co-expressed with HA-tagged Crm1 (red). **c**, As **a**, but the transfected cells were treated with calcium ionophore (iono.) for 30 min. **d**, As **b**, but the transfected cells were first treated with calcium ionophore for 30 min followed by 30 min incubation in the absence of the calcium ionophore. **e**, As **b**, but transfected cells were first treated with calcium ionophore for 30 min and then 100 nM cyclosporin A (CsA) was added.

here are about 50 amino acids long, our results raised the possibility that the binding of calcineurin and Crm1 on NF-AT is mutually exclusive.

If calcineurin and Crm1 compete for sites on NF-AT, activated calcineurin should prevent NF-AT export by Crm1. To test this, we co-expressed NF-AT4(N)ΔZ either with Crm1 alone or with Crm1 and ΔCnA in BHK cells. We found that the expression of ΔCnA abolishes the nuclear export of NF-AT4(N)ΔZ by Crm1, even though calcineurin is not required for importing the NF-AT(N)ΔZ mutant into the nucleus (Fig. 2e). To determine whether the phosphatase activity *per se* of calcineurin is required to override Crm1, we generated a catalytically deficient version of ΔCnA, D130N (Fig. 2c). ΔCnA(D130N), like wild-type ΔCnA, bound NF-AT4ΔZ (Fig. 2d) and resulted in the nuclear localization of NF-AT4(N)ΔZ, despite the presence of overexpressed Crm1 (Fig. 2e). These data indicate that the binding of calcineurin, and not its activity, blocks the nuclear exporting activity of Crm1 towards NF-AT4.

Calcium is known to trigger a conformational change in calcineurin that both activates this phosphatase and enhances its affinity for NF-AT^{1,14,17}. If calcineurin simultaneously promotes NF-AT's nuclear import and suppresses its nuclear export, as suggested by our analysis, then the export of NF-ATs should be affected by calcium signalling *in vivo*. To address this issue, we followed the shuttling dynamics of the GFP-NF-AT4(N)ΔZ mutant as a function of calcium signalling and endogenous calcineurin activity in BHK cells. The GFP-NF-AT4(N)ΔZ mutant is particularly informative because it lacks the NLS-masking Z domain, and is therefore nuclear even in the absence of calcium signalling and calcineurin activation (Fig. 3a). This NF-AT mutant was efficiently excluded from the nucleus by overexpressed Crm1 (Fig. 3b). However, the activation of endogenous calcineurin, through a calcium ionophore, overrode the Crm1 effect and restored the nuclear localization of GFP-NF-AT4(N)ΔZ (Fig. 3c). The antagonistic effect of calcium ionophore towards Crm1 was reversible, as GFP-NF-AT4(N)ΔZ was rapidly exported from the nucleus when calcium ionophore was removed (Fig. 3d). This ability of calcium ionophore to override the export functions of Crm1 on NF-AT4(N)ΔZ was probably due to calcium-triggered activation of endogenous calcineurin, as cyclosporin A, a calcineurin inhibitor that blocks NF-AT–calcineurin interactions¹, restored the exporting activity of Crm1 towards NF-AT in these cells (Fig. 3e). We also noted that when copies of the NES derived from HIV Rev were attached to the N terminus of NF-AT4(N)ΔZ, the export of NF-AT4(N)ΔZ by Crm1 was not longer affected by calcineurin activation (see Fig. H in Supplementary Information). Together, these data demonstrate that the export of NF-AT is subject to strict control by calcium

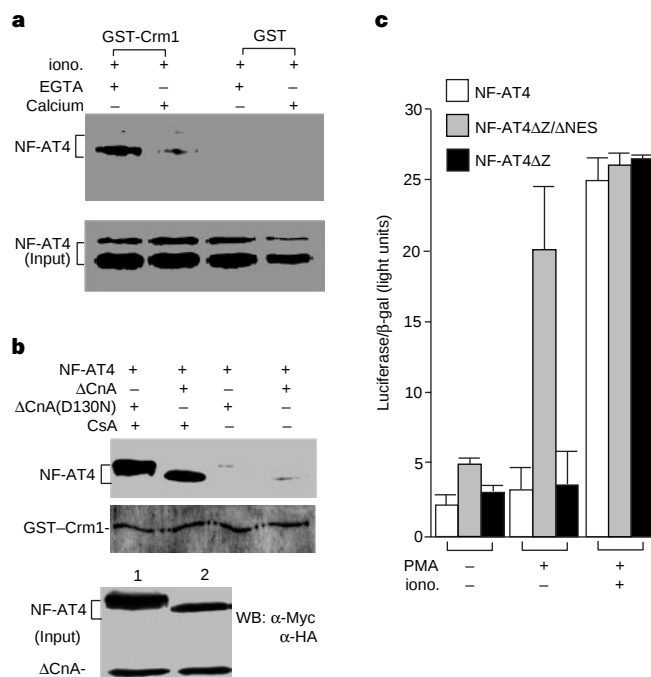


Figure 4 Direct competition between calcineurin and Crm1. **a**, The binding of endogenous calcineurin impairs Crm1/NF-AT associations. NF-AT4-expressing BHK cells were treated with calcium ionophore for 30 min, and then lysed in buffer containing either calcium or EGTA. Immobilized GST–Crm1 was incubated with the cell lysates, and bound NF-AT4 was detected by immunoblotting with the anti-Myc antibody. Input amount of NF-AT is shown in the bottom panel. **b**, Calcineurin binding abolishes Crm1/NF-AT association. Lysates of cells co-expressing NF-AT4 and either ΔCnA or ΔCnA(D130N) were assayed by Crm1 affinity chromatography with and without cyclosporin A (CsA) for NF-AT binding. The amount of calcineurin and NF-AT in cell lysates are shown in the bottom panel (lane 1, lysate co-expressing NF-AT and catalytically deficient ΔCnA(D130N); lane 2, lysate co-expressing NF-AT and wild-type ΔCnA). **c**, Calcineurin binding is required for NF-AT transcriptional activity. BHK cells were co-transfected with NF-AT4 mutants and an NF-AT–luciferase reporter gene, and transcriptional activity was determined as a function of nuclear import and export signals, calcineurin activity (iono.) and the activation of AP1 (PMA).

signalling, and that this control is mediated by endogenous calcineurin, which masks NES sequences on NF-AT.

To analyse further the competition between Crm1 and calcineurin for NF-AT binding, we investigated whether, during calcium signalling, calcineurin binding impairs the association between NF-AT and Crm1. NF-AT4 was transiently expressed in BHK cells, and binding to endogenous calcineurin was induced by calcium ionophore treatment. Lysates of these cells were made under conditions that either retained NF-AT4–calcineurin interactions (1 mM calcium) or abolished them (1 mM EGTA), and they were then absorbed onto a GST–Crm1 affinity matrix. In this assay, GST–Crm1 showed good affinity for NF-AT4 in the absence of calcineurin binding, but poor interactions with NF-AT bound to calcineurin (Fig. 4a).

We further examined NF-AT4/GST–Crm1 association in the presence of overexpressed calcineurin. NF-AT4 was co-expressed with either ΔCnA or catalytically deficient ΔCnA(D130N), both of which bind NF-AT4. Cell lysates containing NF-AT4/ΔCnA complexes were then incubated with GST–Crm1. As shown in Fig. 4b, co-expression of either ΔCnA or ΔCnA(D130N) abolished the interaction between NF-AT4 and GST–Crm1. Additionally, interactions between NF-AT4 and GST–Crm1 were restored when cell lysates were pretreated with cyclosporin A, which disrupts calcineurin binding to NF-AT (Fig. 4b). Together, these data support the

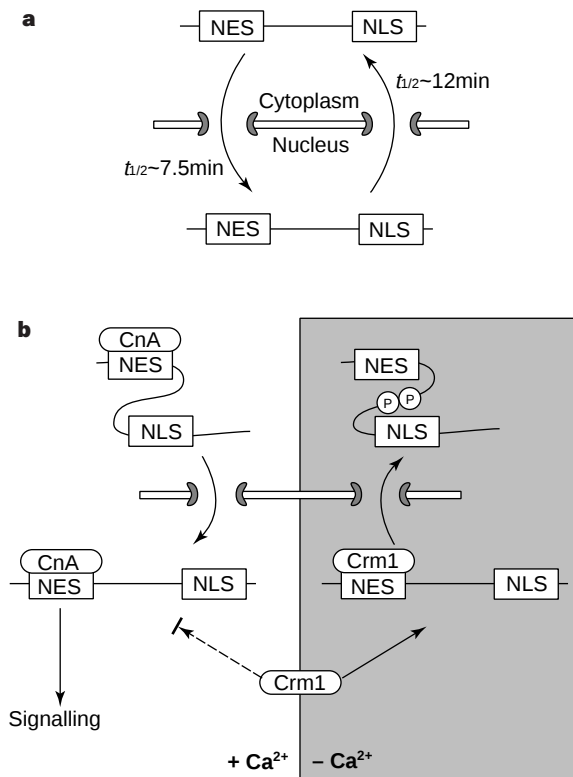


Figure 5 NF-AT shuttling and transcriptional activity. **a**, Scheme for futile nucleocytoplasmic shuttling of a transcription factor bearing unmasked nuclear import and export signal sequences. The transcription factor is subject to constant transport and no transcription signal is generated. **b**, Scheme for rectified nucleocytoplasmic shuttling of NF-ATs. The calcium-activated phosphatase calcineurin binds to NF-AT during calcium signalling and dephosphorylates the NLS-masking domain, resulting in the nuclear import of NF-AT. Calcineurin is co-transported with NF-AT to the nucleus, where it continues to bind to NF-AT by site containing NESs. The exportin Crm1 cannot bind or export NF-AT until calcium signalling ends and calcineurin dissociates from NF-AT. In the absence of calcineurin activity, NF-AT kinases rephosphorylate the NLS mask on NF-AT, further ensuring that this molecule remains in the cytoplasm.

notion of competition between Crm1 and calcineurin for NF-AT binding, and indicate that calcineurin binds NF-AT4 with higher affinity than does Crm1.

When it has moved to the nucleus by activated calcineurin, NF-AT binds DNA through its Rel homology domain and transactivates cytokine genes in cooperation with AP-1 transcription factors¹⁷. To determine whether the suppression of NF-AT export by calcineurin is important for NF-AT's transactivation functions, we examined the transcriptional activity of the NF-AT4ΔZ mutant. Because of its predominant nuclear localization, NF-AT4ΔZ would be expected to be transcriptionally active in the absence of calcium signalling. On the contrary, NF-AT4ΔZ proved to have the same baseline activity as cytoplasmic, wild-type NF-AT4 (Fig. 4c). Calcium ionophore-induced calcineurin binding restored NF-AT4ΔZ to an efficient transcriptional activator (Fig. 4c). These data indicate that, despite its nuclear localization, NF-AT4ΔZ could not form effective transcriptional complexes. Activated calcineurin apparently rescued the transcriptional activity of NF-AT4ΔZ, not by promoting its import, but by suppressing its export. We tested this possibility by examining the transcriptional activity of NF-AT4ΔZ/ΔNES1/ΔNES2, and found that it was transcriptionally active in the absence of calcium signalling (Fig. 4c). It is thus clear that suppression of the NF-AT nuclear export process through calcineurin binding is a prerequisite to the formation of effective NF-AT transcriptional complexes.

The present study describes how NF-ATs, and presumably other

cytoplasmic transcription factors, can signal in the nucleus while being susceptible to the antagonistic processes of nuclear import and export. In the absence of a precise mechanism for regulating export signals, a process of futile cycling, whereby nuclear NF-AT is engaged by components of the nuclear export mechanism, would prevent the formation of functional NF-AT transcription complexes (Fig. 5a). We show that the NF-ATs avoid futile cycling by coupling the activation of import signals to the suppression of export signals. Thus, the binding of calcineurin, the phosphatase responsible for unmasking the nuclear import signals on NF-AT, simultaneously masks the export signals on NF-AT during calcium signalling (Fig. 5b). The model developed here provides a mechanistic explanation for the graded and calcium-dependent signalling by NF-AT, as well as a conceptual basis for trafficking strategies used by other shuttling transduction molecules. □

Methods

Cell lines. HeLa and BHK (baby hamster kidney) cells were grown and transfected essentially as described¹. The calcium ionophore A23187 (Calbiochem), leptomycin B and cyclosporin A (Calbiochem) were used at final concentrations of 1 μM, 20 ng ml⁻¹, and 100 nM, respectively.

Plasmid constructs. All mammalian expression vectors were constructed by cloning complementary DNA inserts into pcDNA3 previously altered by the insertion of a 200-base-pair lamin 5'-untranslated sequence followed by an initiation codon and sequences encoding the Myc, haemagglutinin (HA) or GFP tags^{1,19}. Mutations were made by PCR methods^{20,21}.

Protein binding assays. For analysis of Crm1 binding, GST–NF-AT4(N) fusion proteins were expressed in bacteria and purified by glutathione–Sephadex affinity chromatography. BHK cells overexpressing HA-tagged Crm1 were lysed and equal volumes of cell lysate were incubated with GST–NF-AT4(N) fusion proteins at 4 °C for 30 min. GST fusion proteins were then immobilized on glutathione–Sephadex (Pharmacia) and washed extensively with lysis buffer. The proteins were fractionated by SDS–PAGE, transferred to PVDF (Millipore) membranes and immunoblotted with the anti-HA antibody. To demonstrate the competition between Crm1 and calcineurin, full-length Crm1 was expressed as a GST fusion protein in bacteria and purified by glutathione–Sephadex affinity chromatography. GST–Crm1 immobilized on glutathione–Sephadex was incubated with lysates containing NF-AT4 and ΔCnA at 4 °C for 30 min. NF-AT4 proteins bound to GST–Crm1 were detected by immunoblotting with anti-Myc monoclonal antibody. For co-immunoprecipitation, NF-AT4 protein was absorbed with the anti-Myc antibody and ΔCnA was detected by immunoblotting with an anti-HA polyclonal antibody.

Heterokaryon nuclear export assay. The assay was performed as described previously¹³, with some modifications. Briefly, an NES–GFP–NLS fusion protein was made by attaching the NES sequence of Rev²² to the N terminus of GFP and simian virus 40 NLS to the C terminus of GFP. The NES–GFP–NLS expression vector was expressed in HeLa cells. Myc-tagged NF-AT4(N) mutants were expressed in BHK cells. Heterokaryons were identified as isolated pairs of heterotypic nuclei both containing nuclear GFP.

Received 19 November 1998; accepted 22 January 1999.

- Shibasaki, F., Price, E. R., Milan, D. & McKeon, F. Role of kinases and the phosphatase calcineurin in the nuclear shuttling of transcription factor NF-AT4. *Nature* **382**, 370–373 (1996).
- Zhu, J. *et al.* Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEK1. *Cell* **93**, 851–861 (1998).
- Liu, J. *et al.* Calcineurin-dependent nuclear translocation of a murine transcription factor NFATx: molecular cloning and functional characterization. *Mol. Biol. Cell* **8**, 157–170 (1997).
- Beals, C. R., Clippstone, N. A., Ho, S. N. & Crabtree, G. R. Nuclear localization of NF-ATc by a calcineurin-dependent, cyclosporin-sensitive intramolecular interaction. *Genes Dev.* **11**, 824–834 (1997).
- Masuda, E. S. *et al.* Control of NFATx1 nuclear translocation by a calcineurin-regulated inhibitory domain. *Mol. Cell. Biol.* **17**, 2066–2075 (1997).
- Fornerod, M. *et al.* The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/ Nup214 and a novel nuclear pore component Nup88. *EMBO J.* **16**, 807–816 (1997).
- Fukuda, M. *et al.* CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390**, 308–311 (1997).
- Stade, K., Ford, C. S., Guthrie, C. & Weiss, K. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041–1050 (1997).
- Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, J. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
- Wolff, B., Sanglier, J. J. & Wang, Y. Leptomycin B is an inhibitor of nuclear export: inhibition of nucleocytoplasmic translocation of the human immunodeficiency virus type 1 (HIV-1) Rev protein and Rev-dependent mRNA. *Chem. Biol.* **4**, 139–147 (1997).

11. Nishi, K. *et al.* Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression. *J. Biol. Chem.* **269**, 6320–6324 (1994).
12. Ossareh-Nazari, B., Bachelier, F. & Dargemont, C. Evidence for a role of CRM1 in signal-mediated nuclear protein export. *Science* **278**, 141–144 (1997).
13. Michael, W. M., Eder, P. S. & Dreyfuss, G. The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. *EMBO J.* **16**, 3587–3598 (1997).
14. Luo, C. *et al.* Interaction of calcineurin with a domain of the transcription factor NFAT1 that controls nuclear import. *Proc. Natl Acad. Sci. USA* **93**, 8907–8912 (1996).
15. Loh, C. *et al.* Calcineurin binds the transcription factor NFAT1 and reversibly regulates its activity. *J. Biol. Chem.* **271**, 10884–10891 (1996).
16. Aramburu, J. *et al.* Selective inhibition of NFAT activation by a peptide spanning the calcineurin targeting site of NFAT. *Mol. Cell* **1**, 627–637 (1998).
17. Jain, J. *et al.* The T-cell transcription factor NFATp is a substrate for calcineurin and interacts with Fos and Jun. *Nature* **365**, 352–355 (1993).
18. Hoey, T., Sun, Y. L., Williamson, K. & Xu, X. Isolation of two new members of the NF-AT gene family and functional characterization of the NF-AT proteins. *Immunity* **2**, 461–472 (1995).
19. Taylor, S. S. & McKeon, F. Kinetochores localization of murine Bub1 is required for normal mitotic timing and checkpoint response to spindle damage. *Cell* **89**, 727–735 (1997).
20. Costa, G. L., Bauer, J. C., McGowan, B., Angert, M. & Weiner, M. P. Site-directed mutagenesis using a rapid PCR-based method. *Methods Mol. Biol.* **57**, 239–248 (1996).
21. Tessier, D. C. & Thomas, D. Y. PCR-assisted mutagenesis for the site-directed insertion/deletion of large DNA segments. *Methods Mol. Biol.* **57**, 229–237 (1996).
22. Wen, W., Meinkoth, J. L., Tsien, R. Y. & Taylor, S. S. Identification of a signal for rapid export of proteins from the nucleus. *Cell* **82**, 463–473 (1995).
23. Milan, D., Griffith, J., Su, M., Price, E. R. & McKeon, F. The latch region of calcineurin B is involved in both immunosuppressant-immunophilin complex docking and phosphatase activation. *Cell* **79**, 437–447 (1994).

Supplementary information is available on Nature's World Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank A. Yang and T. Rapoport for helpful discussions and comments on this manuscript, and G. Grosveld, R. Reed and B. Wolff for reagents used in this study. This work was supported by grants from the NIH and the American Cancer Society.

Correspondence and requests for materials should be addressed to F.M. (e-mail: frank_mckeeon@hms.harvard.edu).

Genetically modified

This week's selection of recent introductions has been slanted towards genetics, genomics and genes in general. Beads that bind DNA, membranes that capture RNA and several gene-expression systems feature.

HydroShear

From GeneMachines www.genemachines.com

Generating random DNA fragments

The HydroShear is a simple method for generating random DNA fragments within a user-controlled size distribution. This stand-alone instrument uses hydrodynamic shearing to fragment virtually any source of DNA in any concentration at volumes as small as 100 μ l. By avoiding sample dilution, the HydroShear prevents decreases in concentration before cloning and a consistent DNA fragment size eliminates gel purification and concomitant losses. GeneMachines also claim it's quick: load DNA, run the system and wash, all in about 15 minutes, they say.

Reader Service No. 100

Acrydite

From Mosaic Technologies www.mostek.com

Detect mutants by oligonucleotide modification chemistry

Acrydite has been developed to achieve single-step DNA-attachment and high-speed DNA-hybridization. With Mosaic's Acrydite chemistry, polyacrylamide gels can also be used to analyse nucleic-acid sequences. Nucleic-acid probes are immobilized into the gel matrix, allowing DNA to be attached to any surface in seconds using familiar laboratory methods. The entire hybridization process can take less than 30 minutes.

Reader Service No. 101

Taconic *rag-2* knockout mice

From Taconic www.taconic.com

Mice lacking mature T and B lymphocytes

These mice are unable to initiate V(D)J rearrangement. Taconic's *rag-2* knockout line is maintained on several different inbred mouse backgrounds for specific research needs. The *rag-2* knockout was originally created on the embryonic stem-cell line from the 129/SvEv inbred mouse. On the congenic B6.SJL background, *rag-2* knockout is useful for cell identification in immunological adoptive transfer experiments. The *rag-2* mouse is also available on

fully inbred C3H (H2K) or BALB/c (H2d) backgrounds.

Reader Service No. 102

Low Fluorescence Assay Plates

From Hitachi Genetic Systems

www.hitachi-soft.com

96-well assay plates

These polystyrene flat-bottomed plates have black-pigmented, opaque walls to prevent well-to-well crosstalk and are optically clear for the passage of laser light. Plate bottoms are 60% thinner than conventional polystyrene plates, resulting in lower background fluorescence. Each plate has a total well volume of 360 μ l. 96-well assay plates are available in packs of 25. Plates are sterile, non-treated and without lids and have been tested for low background fluorescence and minimal light scattering.

Reader Service No. 103

Super Agarose

From Molecular Research Center

www.mrcgene.com

Aqueous Agarose suspension

Super Agarose is designed to achieve maximum electrophoretic separation of DNA and RNA. The company claim it is the first agarose available as an aqueous suspension. Weighing of an agarose powder is no longer necessary. Super Agarose is recommended for Southern and northern analysis, PCR and other molecular biology applications.

Reader Service No. 104

DNA synthesis

From Beckman Coulter

www.beckmancoulter.com

Bulk-format DNA synthesis reagents

Beckman Coulter, Inc. now produce bulk packaging formats for DNA synthesis reagents to meet the needs of oligo production facilities. Both derivatized bulk CPG and phosphoramidites can now be packaged



Bulk buy: DNA synthesis reagents by the gram.

in large volumes of one-gram increments to meet manufacturing requirements. In addition, Beckman Coulter can provide liquid reagents in bottles fitting most high-throughput DNA synthesis instruments.

Reader Service No. 105

PCR/ DNA amplification

MicroLYSIS

From Microzone

www.microzone.co.uk

A reagent that allows PCR direct from cells

This new reagent is intended to eliminate the need to prepare DNA samples prior to PCR. By mixing cells with 20 μ l MicroLYSIS reagent and carrying out a simple program on the thermal cycler, cells are lysed and the solution can be added direct to the PCR mix. MicroLYSIS has been tested on bacteria, yeasts, blood, mouth swabs and other mammalian cells. Recommended applications include mass screening of samples and use with PCR robots.

Reader Service No. 106

GenePrint SilverSTR III

From Promega

www.promega.com

Amplification at three loci

The GenePrint SilverSTR III allows for the simultaneous single-tube amplification of three polymorphic STR loci with non-overlapping allele size ranges: D16S539, D7S820 and D13S317. The amplified STR fragments can then be separated by polyacrylamide gel electrophoresis and detected by silver staining.

Reader Service No. 107

Gene expression

MicroMax

From NEN

www.nenlifesci.com

A glass-based 'DNA chip' pre-formatted for differential gene-expression analysis

The MicroMax Human cDNA cDNA Microarray System I contains two glass microarray slides, each pre-spotted with 2,400 known genes. Most of the genes on the microarray, the first product in the MicroMax range, are from human cDNA libraries developed by AlphaGene, Inc., and some plant control genes are also included. The genes are characterized by a Prosite 'motif' search and are identifiable via NEN's website, which links to NIH GenBank. Detection and analysis is performed on a microarray scanner, such as the ScanArray 3000 (General Scanning, Watertown, MA), or by NEN's custom slide reading and image analysis service.

Reader Service No. 108



Lose your rag: a knock-out in several strains.

Panorama Gene Arrays

From Genosys

www.genosys.com

A tool for genome-wide analysis

Panorama *E. coli* gene arrays, from Genosys Biotechnologies, provide a rapid way to determine global gene expression during cellular responses to developmental cues or external stimuli. The arrays contain 4,290 PCR-amplified open reading frames representing all *E. coli* protein-coding genes. Available separately are cDNA labelling primers optimized for *E. coli*.

Reader Service No. 109**GFP variants**

From Clontech

www.clontech.co.uk

Rapid-turnover variants of GFP

These are green fluorescent protein variants for use in quantitative reporter assays and kinetic studies. Expression can be rapidly visualized by fluorescence microscopy or analysed by flow cytometry. No additional substrate or cofactors are required, allowing destabilized variants to be used in tracking dynamic cellular processes *in vivo* and in real time. For mammalian expression, destabilized enhanced GFP variants are available with half-lives of one, two and four hours. The bacterial destabilized GFP Vector Set contains three destabilized GFP variants with half-lives of 40, 60 and 110 minutes.

Reader Service No. 110**CytoBlast**

From Packard

www.packardinst.com

FRET-Based Gene Reporter System

Packard Instrument Company offers a new fluorescence resonance energy transfer (FRET)-based system for the study of *cis*-acting sequences and *trans*-acting factors. CytoBlast relies on an enzyme-substrate reaction to monitor gene expression. A putative promoter element powers production of β -lactamase. The level of enzyme production is detectable by addition of the CytoBlast CCF2/AM substrate ester. Each intact substrate molecule undergoes FRET and emits in the green spectrum when excited. β -lactamase cleaves the molecule and disrupts FRET, shifting emission towards the blue. This emission change is directly related to the promoter activity and allows for single-cell, real-time gene-expression reporting.

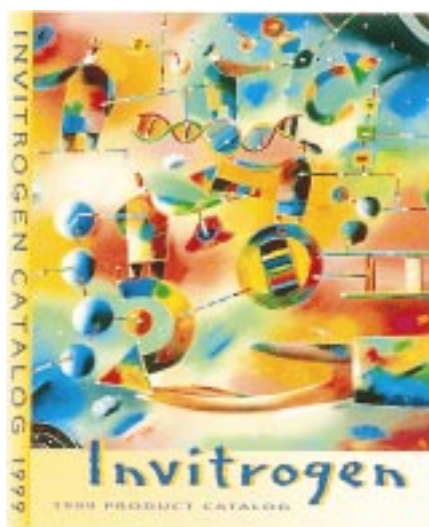
Reader Service No. 111**Blasticidin**

From Invitrogen

www.invitrogen.com

An antibiotic isolated from Streptomyces griseochromogenes

Blasticidin is a potent translational inhibitor in both prokaryotic and eukaryotic cells and is marketed as a tool for researchers express-

**All the fun of the lab: Invitrogen's 1999 catalogue.**

ing genes in mammalian cells. *E. coli* strains are generally sensitive to 50 μ g per ml while mammalian cells are sensitive to 1–2.5 μ g per ml. Cell death occurs rapidly at these low concentrations. Resistance to blasticidin is conferred by the *bsd* gene from *Aspergillus terreus*. The mammalian expression vectors pcDNA6/V5-His A, B and C provide the *bsd* gene and allow clonal isolation of blasticidin-resistant mammalian cell lines in as little as 7 days post-transfection. pcDNA6 vectors utilize blasticidin for selection. Invitrogen's 1999 catalogue, listing 300 products for gene cloning and expression, is now available, fax (760) 603-7201.

Reader Service No. 112**Genomic DNA****Wizard Genomic DNA Purification Kit**

From Promega UK

www.euro.promega.com

A large-scale kit for purification of high-molecular-weight DNA

Developed for the purification of high-molecular-weight DNA from 10- to 300-ml blood samples, this system contains the reagents for 100 10-ml isolations and the procedure takes only 45 minutes. Reported yields are consistently 500 μ g per 10-ml sample and the purified DNA is suitable for amplification, restriction-endonuclease digestion and membrane hybridizations.

Reader Service No. 113**PASSPORT**

From Amersham Pharmacia

www.apbiotech.com

Fast genetic mutation and polymorphism analysis

PASSPORT kits use enzymatic mutation detection (EMD) technology to perform mutation scanning and polymorphism screening for unknown mutations which may be associated with the cause, diagnosis or treatment of a disease. The kits can be

used on segments of DNA as large as several kilobase pairs in length. EMD technology makes use of resolvases to recognize mismatches in DNA and cleave the DNA close to the mismatched site. The resulting fragments are then simply separated and evaluated by electrophoresis.

Reader Service No. 114**Phase Lock Gel**

From Flowgen

www.flowgen.co.uk

Enhanced nucleic-acid recovery

Flowgen's Phase Lock Gel can be used with any protocol requiring phenol or chloroform extraction for the isolation of genomic, plasmid or phage DNA, total RNA, DNA from cells and DNA from mouse tails. On centrifugation, Phase Lock Gel migrates to form a tight seal between the phases of an aqueous/organic extraction. The organic phase and the interface material are effectively trapped by this barrier, allowing easy transfer of the entire aqueous phase by simple pouring or pipetting. The entire aqueous phase is collected, so nucleic-acid recoveries can be enhanced. Flowgen claim a 20–30%

ADVERTISEMENTS

**BMA BIOMEDICALS AG**

Your source for unique antibodies:

MCP-1

Scavenger Receptor CD163

Endothelial and Macrophage markers

Murine myeloid cell markers

hMRP8/14 (Calprotectin) ELISA

hIL-1 receptor antagonist ELISA

Call for your free catalogue:

++41 61 811 6222 (phone)

++41 61 811 6006 (fax)

READER ENQUIRY NO. 1

Check out
our homepage at
<http://www.resgen.com>

BACs**YACs****PACs****cDNAs**

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.**U.S. or Canada 800-533-4363****U.K. 0-800-89-1393****FAX 256-536-9016**

READER ENQUIRY NO. 30

new on the market

efficiency increase with Phase Lock Gel.
Reader Service No. 115

RNA detector

From KPL www.kpl.com
A non-radioactive RNA detection system

This new non-isotopic detection system offers sensitivity equivalent to ^{32}P in less time, and with only 1 µg of RNA on the gel. Biotinylated probes are stable for months, and can be reused many times. The system supplies hybridization buffer, block and wash solutions, AP streptavidin and chemiluminescent substrate for consistency and to minimize preparation time.

Reader Service No. 116

Dynabeads

From Dynal www.dynal.no
PCR-ready genomic DNA from blood

This system from Dynal has been developed for high-throughput automation of DNA isolation from whole blood for downstream PCR applications. It combines a reagent kit (Dynabeads DNA DIRECT Auto96) with an automated magnet station (Dynal MPC-auto36) interfaced with a control PC and any liquid-handling robot. The system han-

dles blood volumes of 10–30 µl in a 96-well format and processes 96 samples in 80 minutes. DNA yield is typically between 600 ng and 1 µg from 30 µl blood, enough for up to 50 PCR reactions at 20 ng per amplification.

Reader Service No. 117

RNA

NANO BIO

From Patentic www.patentic.fi
Air purification for gene laboratories and other sensitive areas

NANO BIO is a novel system for cleaning air, relying on air flow and ducting rather than filters. The system runs air through 1.5-m ducting and the unit can be conveniently placed in a room next to the wall.

Reader Service No. 118

RNA oligonucleotides

From Dharmacon Research
www.dharmacon.com

RNA oligonucleotides in 4 days

Dharmacon's aim with this new system is to make RNA synthesis as reliable and convenient as DNA synthesis has become. The chemistry employs a new 5'-silyl protecting group in conjunction with an acid-labile 2'-

orthester protecting group (2'-ACE). The 2'-ACE groups are rapidly (in less than 30 minutes) and completely removed under mild conditions in common aqueous buffers. Standard RNA oligos can be shipped in four business days.

Reader Service No. 119

mRNA capture membranes

From Trevigen www.trevigen.com
Rapid and specific capture of mRNA from lysed cell or tissue samples

Trevigen's simple filter discs are added to blood or other biological samples together with lysis solution. Cell lysis and mRNA capture take place in a single tube containing a strong chaotropic salt that inactivates endogenous RNases. Captured mRNA is quickly removed from the nucleases that threaten the quality of the RNA preparation producing mRNA ready for RT-PCR — which can even be performed directly on the filter.

Reader Service No. 120

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

RECOMBINANT SERVICES

ProteinGenesys
Proteins for Sciences of Life

We Can Deliver Purified, Native & Glycosylated, You Just Name the Protein, We Do The Rest

Proteins for Research, Applied Protein Chemistry, Drug Discovery, Genomics, Immunochromatography and More

www.protingenesys.com tel: +41 61/811 3357, 2803 fax

READER ENQUIRY NO.

5

The Mouse Kit™
for Gene Targeting in ES Cells

- ✓ Complete with embryonic stem (ES) cells, feeder cells and essential reagents (including FBS) for engineering one knockout mouse strain
- ✓ Low-passage AB2.2-Prime ES Cells® (129/SvEv)
- ✓ Uniform, high-quality ESQ feeder cells (mitotically inactivated STO cells) engineered for *neo* and *puro* resistance
- ✓ Prequalification of ESQ feeder cells and reagents for use with AB2.2-Prime ES Cells

*Manufactured by Lexicon Genetics Incorporated

Fax: (619) 535-0071

Tel: (619) 535-5400 or (800) 424-5444

www.stratagene.com

READER ENQUIRY NO.

6

contact: info@TIB-MOLBIOL.de

TIB
MOLBIOL

Molecular Beacons...

...and other fluorescent-labelled oligonucleotide probes for advanced nucleic acid detection. Dark quencher for all dual-labelled probes available.

Custom Synthesis of Molecular Beacons for research use under licence of The Public Health Research Institute, New York.

READER ENQUIRY NO.

8

Detect residual DNA with Wako's DNA Extractor Kit or Isolate Genomic DNA with DNA Extractor WB Kit

Both use the Nal Method
No Phenol No Chloroform

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO.

39

Have you ever really believed that you are merely an intelligent animal that evolved by chance on a planet that exists by chance in a happenstance Big Bang universe?

If not, then read:

The Honesties of Einstein
The Existence Mechanisms of the Steady State Universe
@
www.jrlpub.com

Published by
Llano Research Laboratories

READER ENQUIRY NO.

310

Physics grapples with its image problem

Fewer young people are wanting to study physics but universities are fighting back by designing more relevant courses. Graduates in the subject find plenty of openings in industry but a shortage of posts in the academic world.

Helen Gavaghan

Newspapers are awash with stories of cloning, biotechnology and genetic modification. Medical breakthroughs are reported daily. Governments, arguing that the biological sciences are an important basis for future economic well-being, and for people's health, have no difficulty justifying the decision to hand large chunks of taxpayers' money to biological research. In the United States, for example, the National Institutes of Health won an astonishing 15 per cent increase in its budget for 1999. In industry, the picture is equally rosy, with biotechnology growing and pharmaceutical company shares holding their own.

In short, it's cool to be a biologist, and the justification for a biology career choice is clear to anyone who can read. You can talk about biology at parties, and people will look you in the eye and say, "That must be really interesting."

What of physics? Here is an endeavour tackling the deepest mysteries of matter and energy as well as providing the underpinning for much of science, including biology. The discipline plumbs imaginative depths, its practitioners never knowing what greater depths may lie below. It brushes the fringes of philosophy and the unknown. Yet when the UK Institute of Physics asked academically bright teenagers why they were not studying physics after the age of 16, some gave the unbelievable response that they were interested in ideas and so would rather not study physics, thank you very much.

This perception of physics as dull and irrelevant may partly explain why fewer UK students are studying it after the age of 16 and why applications for undergraduate courses are falling. Around 12,000 fewer students took A-level physics (an exam that students sit when they are 18 and preparing for university) in 1998 than the nearly 47,000 who sat the exam in 1989.

In the United States, by contrast, the number studying physics at school has risen from 20 to 28 per cent over the past decade. Yet all is not rosy in the United States. In 1997, US institutions granted 3,826 bachelors degrees in physics, the lowest number in 40 years and a 24 per cent decrease since 1989. The decrease, mirrored in other physical sciences and in engineering, is partly attributable to the fact that there are fewer 18- to 24-year-olds now than 10 years ago.

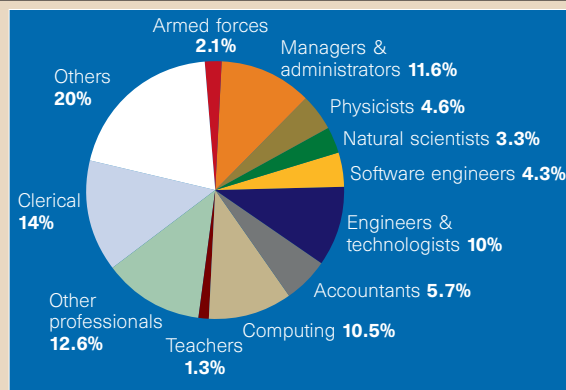
Demographics is not the whole story, however. Students who have prepared to study physics at university often change sub-

Bright prospects and high salaries

One of the most visited pages at the American Association of Science's website is called salary survey. Actually it is about more than salaries, showing how PhDs from 14 disciplines compare in the job market. Physicists come out well in the survey, having nearly the lowest unemployment level at 2 per cent, as well as comparatively high salaries.

The median salary for a PhD physicist in business/industry is US\$62,000; the median for government work is \$63,000. This compares with \$72,500 and nearly \$55,000, respectively, for computer scientists in business/industry and government.

Nevertheless, economists beat both physicists and computer scientists, with a median starting salary of \$73,500 in business/industry, a figure which might explain the



Options: areas of work of UK physics graduates surveyed in 1996.

attraction some physicists feel for Wall Street and the City of London (see page 267).

In the United Kingdom, the Institute of Physics carries out a salary survey every two years. In 1998, the median salary across all sectors and levels of qualification was £27,850 (\$45,000), an increase of 5 per cent compared with 1996. The highest-paid jobs were in industry. The median salary in the

chemical and petrochemical industry, for example, was £37,500. In telecommunications and electronics, the median was £35,500.

Academic physicists, however, will have to find the intellectual challenge enough in itself – their salaries being below those in industry and other fields. Median salaries for those in universities were £28,000 and for those in research laboratories £26,173.

jects in their first year, says Alvin Saperstein, a physics professor at Wayne State University in Michigan. It is as though US students do not become as disillusioned about physics as their British counterparts until their first year at university.

Saperstein thinks falling numbers of physics students is the universities' fault for not paying enough attention to teaching. The consequences of that attitude, he argues, are now coming home to roost because, if there is less teaching to do, retiring members of staff may not be replaced. Top-heavy departments such as Saperstein's — 30 faculty members for a department turning out two or three physics majors per year — will be particularly vulnerable.

There are more faculty members in the United States over the age of 60 than under the age of 40, so the issue of what happens as staff retire will come to a head during the next five years or so. "If the decline in enrolments bottoms out at 1997 levels, it will be OK," says Roman Czujko of the American Institute of

Physics (AIP). "But if it continues there will be pressure on faculty." What happens at the undergraduate level will set the stage for academic physics for the next decade.

A decline in the number of faculty positions for physics would make the task of finding a full-time, permanent job in academic research even more crushingly difficult than it is today. The current difficulty arises because the number of new physics PhDs has risen sharply in recent years. Although the number of postdocs on the market has risen, the number of faculty positions becoming vacant each year has stayed the same.

In the early 1990s, the competition increased further when corporate mergers and downsizing put many mid-career physicists on the market. The end of the Cold War also made internationally renowned eastern European physicists available for work. For the US postdoc class of 1993, the situation was abysmal, and remained that way up to about 1996, easing off now to merely 'pretty bad'.

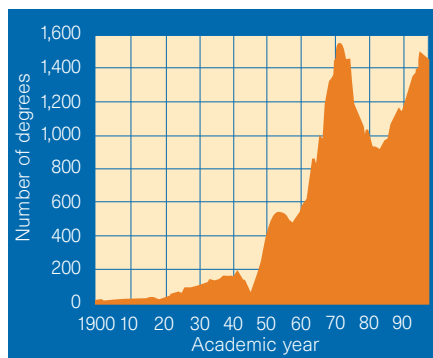
SOURCE: HESA

Rick Moyer, a research physicist at the University of California, San Diego, believes it is unethical for institutions to continue to churn out new PhDs trained for academic research, most of whom will be hoping for a tenure-track position, when such positions are like gold dust. Moyer argues that PhD training could perhaps include business studies to prepare physicists for jobs in industry.

Worldwide problem

In France too, where 2,200 physics PhDs were awarded last year, there is a need for training to reflect the reality that most PhDs have to find work in industry, says Marc Joucla, director of the Association Barnard Gregory. ABG is a non-profit group offering seminars to doctoral students from different disciplines to help them to acquire a wider professional perspective than academic research alone.

Helmut Krauth of the German Physical Society, an industrialist with a metals company, Vacuum Schmelze, says that German industry is in general happy with the quality of physicists with PhD and diploma education. Nevertheless, he says, "professors should encourage students to get work experience during the summer". Such training



Education peaks: US physics PhDs rose in the 1980s, outstripping the number of faculty posts.

will set them up for work in industry, he says.

Moyer's argument for commercially based training had more force a few years ago than it does today, because companies used to be leaner and more focused. They did not want the expense of taking on PhD physicists who would need training for the job. In those stringent times, the much vaunted ability of the physicist to be flexible and apply skills to a variety of problems was not attractive when a company could employ someone with just the specialist skills they were looking for. Now that the US economy is booming,

industry is once again being creative and expanding, and is attracted by the generalist skills of physicists. Software and telecommunications are particular growth areas.

Even though physicists are once again in demand, there is still a need to prepare PhDs for the reality of their future careers rather than the dream of success in the academic sphere. Native US physicists with PhDs now have to compete against physicists from elsewhere lured to the United States by companies offering better salaries than available in their own countries but lower than paid to their US counterparts, says Moyer.

Industry beckons

The consequence of a strong economy and the difficulty of getting postdoc positions that could lead to tenure-track jobs are reflected in the latest figures from the AIP. These show that the numbers accepting jobs in US industry have risen during the past decade. Ten years ago, just under half of all PhDs who took permanent jobs went to industry. Now the number is 70 per cent.

Not all of those in industry are happy. "Some are disappointed to be in the private sector because they thought they wanted to be physicists when they grew up," says

Exciting times at the interface with other disciplines

Multidisciplinary research is of increasing importance, according to the US National Science Foundation. "We are seeing excitement at the interface," says Jack Lightbody, executive officer of the foundation's physics division. Lightbody says there is work to be had for physicists in areas such as modelling social and biological systems, and in medical physics.

The UK Engineering and Physical Sciences Research Council (EPSRC) is thinking along similar lines. Its policy document published this month, *Programme Landscape 1999-2000*, sees opportunities "for pioneering research across traditional boundaries". Helmut Krauth of the German Physical Society agrees: "There is a tendency to do more interdisciplinary research," he says.

The EPSRC, which funds university physics research (excluding particle physics and astronomy), plans a series of meetings to help physics cross interdisciplinary boundaries.

One example will focus on modelling and simulation for physicists, people in industry and other academics. Interdisciplinary studies involving physics and environmental science and physics for healthcare will continue to be encouraged.

One new activity should help future job hunters interested in biophysics. The EPSRC plans to explore the opportunities in this area, believing that physics has much to offer the study of protein signalling networks, protein folding, biopolymers and colloids.

The physics and astronomy board of the US National Research Council also recognizes the value of links with biology in these days of shrinking physics budgets and blossoming biology funding (see *Nature* 397, 89; 1999). Explaining its current study into the future of physics, the board says: "To attract ambitious people in the future, it is necessary that physics concern itself with rapidly developing areas such as medicine and biology."

Several US universities have already spotted the intellectual value of promoting collaboration between physics and biology (see *Nature* 397, 3; 1999). Stanford University is investing in a new institute with 50 faculty members (including 10 new posts) that will draw together disciplines as disparate as applied physics and clinical medicine.

A survey conducted last year for the United Kingdom's higher education funding bodies confirmed the pervasiveness of interdisciplinary research in higher education. Three-quarters of researchers are doing interdisciplinary work. The funding agencies commissioned the study as part of their consultation on the quality of the country's methods of assessing university research and apportioning funding. They wanted to know whether the research assessment method inhibits interdisciplinary research.

Preliminary findings are that most researchers are involved in both single discipline and

interdisciplinary research. Of these, 23 per cent believe that the research assessment exercise strongly inhibits interdisciplinary research. Those who spend most of their time on interdisciplinary research have the greatest concern. Heads of academic departments for the most part thought that the assessment exercise had little effect on interdisciplinary research. The survey also found that research assessment panels are less interdisciplinary than the scientists whose work they are evaluating.

The full analysis, by Evaluation Associates, is due to be published soon. It is likely to extend the preliminary recommendations that more scientists with knowledge of interdisciplinary work should be included in research assessment panels. If physics is to continue reaching into new areas, implementation of these recommendations will be important.

The preliminary findings can be found at: <http://www.evaluation.co.uk>

Czujko. He hopes to follow up some of these PhDs to discover their views in a few years' time. "My guess is they will be real happy. These people are earning entry-level salaries of \$60,000 to \$70,000 per year. I want to feel their emotional pain, but..."

Perhaps sympathy should be reserved for the 43 per cent of 1997 physics PhDs who took postdoc positions. A study by the Commission on Professionals in Science and Technology published 18 months ago found that many new PhDs are forced to spend increasing lengths of time in low-paid, temporary, postdoc positions. The survey showed that of PhDs who took postdoc positions in all subjects (not just physics) in 1993, only 13 per cent had achieved tenure-track positions two years later. By 1995, the figure in tenure-track posts was only 16 per cent.

The AIP survey of 1997 PhDs shows that, of those in temporary postdoc positions, three-quarters were in those positions only because they could not find a full-time, permanent post. Because physics faculties are unlikely to increase in size, the best that can be done to minimize the waste of highly trained postdocs in low-paid temporary posts is to try to keep faculties with the same staffing levels, which means attracting physics majors and holding their interest.

Tailor-made courses

One institution with an imaginative response to the problem is Rutgers University in New Jersey, which draws from a blue-collar population. The students are often the first in their family to go to university and are people who are likely to want to get a job after their first

degree rather than going on to do a PhD. This is not because they are not smart enough, but is the result of economic reality. Rutgers is teaching a physics course deliberately structured to this reality. About 800 US institutions offer bachelors degrees in physics, and many are seeking ways to change the curriculum to attract more students.

At masters level, too, there is a need to tailor education to students' realistic job prospects. Surprisingly few masters courses include options of obvious value to business. In the United States there is the additional problem that holders of masters degrees tend to be perceived as failed PhDs rather than as bona-fide holders of a higher degree.

Those who want to study physics to bachelors, masters or doctoral level would be well-advised to question whether the course

Modelling makes the money go round

The esoteric and intriguing world of derivatives trading is attracting increasing numbers of physicists. As yet only a trickle of some tens of people each year are joining the business in the United States and United Kingdom, but nevertheless the financial sector is an emerging job market for physicists.

Derivatives trading can be speculative and purely for profit, even to the extreme of becoming a gambling addiction. Equally, however, this field can satisfy a socially responsible and economically meaningful function by, for example, minimizing the risk of financial institutions. Banks shelter from the unpredictable winds of economic fortune by buying or selling derivatives. This allows banks to sell mortgages to individual customers at affordable levels and stay in business, despite fluctuations in interest rates.

A derivative is a contract whose value derives from the value of an underlying security such as a stock price, interest rate or foreign exchange level. The derivative might be an option to buy a stock at a given price in the future. In other words, the holder of an option can buy a stock for, say, \$150 on a given day in the future. If the value of the stock on that date is \$180, the individual will exercise his or her option and buy the stock at \$150. If the

value is lower, the individual will not buy. Either way, the individual will have paid the derivative trader for the option.

The financial institution selling the derivative must, of course, be able to provide the stock at the value agreed if the customer exercises his or her option, but it clearly does not want to lose money if the stock has soared in value beyond that anticipated. Myron Scholes and Fisher Black, an economist and an applied mathematician working together at the Massachusetts Institute of Technology, came up with a formula in the 1970s for pricing options.

Another class of derivatives is known as futures. The idea behind futures is easiest to grasp in the context of a commodity such as orange juice. The holder of a futures contract agrees to buy orange juice at a given price in the future, and thus the seller of the contract assumes the risk of a possible increase in the price. Unlike in the case of an option, the contract holder cannot walk away from the deal if the price of orange juice declines. The customer pays the dealer for the privilege of being protected from that risk.

Options are the simplest types of derivatives, says Andrew Lesniewski, a former Harvard physicist who made the move to Wall Street in preference to heading a

mathematics department at a US university. Currently, Lesniewski trades exotic derivatives at the French investment bank Paribas.

More complex derivatives are contracts that protect the buyer from unfavourable variations in, say, interest rates or foreign exchange. The buyer could be a mortgage company worried about refinancing when interest rates decline, a US portfolio manager holding foreign stocks in his or her portfolio, or a European company whose balance sheet depends on the US dollar price of crude oil. The derivatives dealer decides how much he or she will charge to assume the risk of the unfavourable market movement in interest rates, foreign exchange, or oil price. "Derivatives can be as customized and as complex as you want," says Lesniewski.

To protect itself, the derivatives trader will have its own constantly changing portfolio of derivatives. The trader needs to hedge its portfolio against future movements of a range of commodities, stocks and financial indicators. It must price its derivatives so that they are good value to the company needing to off-load risk, while remaining viable itself despite the risk it is carrying. Risk management of a portfolio of derivatives held by a financial institution is the key role of a

trader. In the case of interest-rate derivatives, this involves sophisticated mathematical models of the shape and dynamics of the yield curve, or term structure of interest rates.

Hence there is a need for people able to reduce complex systems to their essentials and to describe them mathematically. "The people we recruit must be smart," says Lesniewski. They need a sophisticated knowledge of probability theory, stochastic processes, partial differential equations and numerical analysis. It is a job that physicists, statisticians and mathematicians can do. "What I find challenging," he says, "is that it is a new, dynamic and growing area of the economy. You can do original work, make a significant contribution and see concrete results."

Yet there is a need for caution. Derivatives traders need a realistic knowledge of the role of money. "It's not just that money is the bottom line in this business," says Lesniewski. "Derivative trading is about buying, selling and managing money and predicting how economic factors will influence its behaviour. If you are not comfortable with money, it is not for you."

Further reading:

Hull, J. *Options, Futures, and Other Derivative Securities* (Prentice Hall).

offered suits their needs in the modern world, or is one that suits a faculty's traditional view of what physics should be about.

Declining interest

A similar story of declining interest in physics among university applicants is to be heard in the United Kingdom. The number graduating with a bachelors degree is holding steady at about 2,300 per year (an apparent dramatic dip in physics graduates in 1996 was an artefact resulting from the introduction three years earlier of four-year courses), but the percentage of applicants for physics degrees is down by 10 per cent compared with 1998. This compares with an average 2 per cent decline nationally in all subjects. The decline in the numbers taking up post-graduate positions has different reasons. The minimum stipend for a PhD is £6,500 (\$10,000) per year, not enough to live on (*Nature* 397, 640; 1999).

In an effort to reverse the decline in physics at school level, the UK Institute of Physics has devised a curriculum to make the subject more attractive to 16–18-year-olds. This programme goes beyond classical physics and introduces students to cosmology and particle physics. A few schools will take part in a pilot study starting this September to test the course. If successful, the course will be widely available in September 2000.

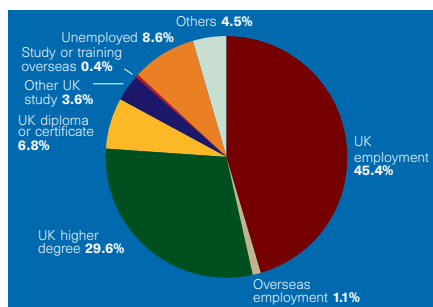
Cold wind blows

Although there are few hard numbers to back up the perception, it seems that the coolness towards physics seen in the United States and the United Kingdom is also being felt across northern Europe. "It's largely hearsay, but throughout northern Europe there are anecdotal stories of declining numbers enrolling to study physics," says John Lewis, treasurer of the European Physical Society. So concerned is the society that in September it is bringing together 50 influential physicists from across Europe, including presidents of 16 national societies, to thrash out a strategy for discovering how much of a problem there is and how it can be dealt with.

Germany is likely to feel the pinch soon. In 1998 for the first time more physics PhDs were awarded than physics diplomas (a degree that is more than a masters but less than a doctorate), says Krauth. In Germany a diploma is a highly regarded qualification awarded after five years of higher education and essential for anyone wishing to earn a doctorate. The decline in diplomas, which has taken everyone by surprise, will mean that Germany's production of PhD physicists will fall in a few years, says Alex Bradshaw, head of the German Physical Society.

Social conscience

The factors causing the declining interest in school or undergraduate physics are complex. In the United States and the United



First destinations: work or further study chosen by recent UK physics graduates surveyed in 1996.

Kingdom, one reason is the unimaginative quality of curricula and teaching methods in first-year university courses or at 16-plus. But there are bigger issues at stake. Krauth says: "In Germany, young people are questioning whether physics and technology have served humanity well."

In the United States, the physics and astronomy board of the National Research Council (the research arm of the National Academy of Sciences) began a decade-long series of studies in 1991 to confront head on the issue of the relevance to society of all branches of physics. Central questions being addressed include: 'How can the intellectual vitality of physics be continued into the next millennium?' and 'What contribution does physics make to the national need?'

In these post-Cold War days the answer to the second question is less obvious to the public, and hence to prospective students, than it used to be. One problem is that there is no physics industry, says Czujko. Chemistry has the chemical industry, biology has biotechnology and pharmaceuticals, geology has mining and oil exploration, but the path from basic physics to industrial usefulness and jobs is not obvious.

Conscious of this low visibility of physics, the physics and astronomy board's first study of a branch of physics, published in 1994,

pointed out that atomic, molecular and optical physics is the underpinning science for 9 per cent of US gross domestic product. During 1999, the board will publish studies of elementary particle physics, condensed matter and materials physics, nuclear physics and gravitational physics. Then will come the third and most difficult stage, an overview of physics, to be published in 2000 and making recommendations about education and attempting to define the place of physics in today's society.

The low profile of physics leaves students with the false impression that job options are limited. Yet, providing that they do not aspire to a full-time, permanent academic research post, this is a false perception. In 1997, only two per cent of US PhD physicists were unemployed after six months. But of the 48 per cent in permanent positions, only 43 per cent were employed in physics. The rest had jobs in engineering, computing, other sciences, business and finance. At Vacuum Schmelze in Germany, Krauth says that 60 physicists out of a total workforce of 2,000 do a full range of jobs from research and development to sales and marketing.

A UK survey by the Higher Education Statistics Agency shows that people with a bachelors degree in physics found jobs in 1996 in business analysis, as test engineers, in sales, data research, the armed services, as systems engineers, analysts and programmers, as teachers and as production, telecommunications and safety engineers. Just under 30 per cent were studying for PhDs or masters degrees.

Perhaps then, despite the gloom in the academic sphere, there is some justification for Czujko's assertion: "I don't think physics is dead. I'm bullish on physics." □

Helen Gavaghan is a freelance science writer at 27 Edge Hay Green, Colden, Hebden Bridge, Yorkshire HX7 7HJ, UK.

e-mail: 101543.30@compuserve.com

Job search starts here

The American Institute of Physics website is the most relevant site for US physicists, containing forums, links and a wealth of statistics, survey data and job advertisements, as well as a section in which to post your resumé. Other useful URLs are those of the American Physical Society and the American Association for the Advancement of Science.

In Europe, start with the European Physical Society site. This connects you with the physics societies in

Europe and gives access to sites like that of the non-profit group Association Barnard Gregory in France, a resource for job-hunting scientists. The UK Institute of Physics is the best funded European organization, and offers the most extensive resources and information about education and careers of the European sites.

For electronic searches, remember the bulletin boards and USENET. If you send unsolicited CVs, remember that they may be

scanned into a database. Make sure the layout is clean and unfussy with no fancy typefaces. Make sure the CV contains the crucial search words that you would want to be identified to pull your CV to the fore. Good luck!

<http://www.aip.org>
<http://www.aps.org>
<http://www.aas.org>
<http://www.iop.org>
<http://epswww.epfl.ch/dir/>
<http://www.physlink.com>
<http://invest-faq.com>